Radical Chemistry for Selective Modification of Native Residues in Proteins



Mateusz Imiołek

Oriel College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2020

Radical Chemistry for Selective Modification of Native Residues in Proteins

Mateusz Imiołek

Oriel College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2020

The chemical modification of proteins has provided a toolbox enabling many experiments in chemical biology. However, new strategies are required to be able to specifically target (hetero)aromatic residues. Fluorine atoms installed in proteins can serve as useful tags, but methods for efficient and selective chemical fluorination are limited. This thesis has explored the modification of native residues in proteins with fluoroalkyl radicals to address the shortage in current methodologies. The study began with the investigation of trifluoromethylation, using biocompatible oxidative generation of CF_3 radicals from a sulfinate precursor, and resulted in a robust protocol for selective modification of tryptophan residues. Proteins modified with this minimally perturbing tag could be readily studied by protein observed ^{19}F NMR. To evaluate possibility of ^{18}F-trifluoromethylation, the automated radiosynthesis of the radical precursor was achieved, but unfortunately, the protein radiolabelling proceeded with a limited efficiency. On the other hand, the study of protein difluoroethylation, using the same radical generation system, revealed an unexpected N-modification of proteins, for which the mechanism was proposed (oxidative hydrolysis of fluoroalkyl sulfinates). To reduce this side reactivity, an improved fluoroalkylation protocol was developed, based on the photoredox activation of heteroaryl sulfones as radical precursors. The platform was used to study reactivity and selectivity trends of various fluoroalkyl radicals, which resulted in the detailed investigation of difluoromethylation. The preference for the formation of unstable difluoromethyl tryptophan products was used for the development of selective C-formylation of proteins, relying on a key step of the denaturative hydrolysis of the CF_2H group. As such, a new method was established for the installation of a masked aldehyde, also serving as a minimal size fluorescent probe. Overall, the studies of radical fluoroalkylation of proteins yielded useful methods for modification of typically solvent inaccessible residues (specifically tryptophan), under-targeted in protein chemistry, with applications in protein observed NMR and bioconjugation.

Contents

Acknowledgements **viii**

Abbreviations **ix**

1 Introduction **1**

1.1 Chemical modification of proteins 2

1.1.1 Modification strategies for engineered proteins 4

1.1.2 Strategies for native residue modification 10

1.2 Radical chemistry for the editing of biomolecules 31

1.2.1 Radicals in protein chemistry 33

1.2.2 General features of fluoroalkyl radicals 38

1.3 Fluorination of proteins 45

1.3.1 Generation strategies 46

1.3.2 Applications 49

1.4 Aims of this thesis 54

2 Selective trifluoromethylation of native residues in proteins **56**

2.1 Introduction 58

2.2 Small molecule and peptide model 59

2.2.1 Reactivity 59
2.2.2 Selectivity 66
2.2.3 Peptidic substrates 68
2.3 Protein trifluoromethylation 71
2.3.1 Myoglobin as a model protein 71
2.3.2 Reaction optimisation 72
2.3.3 Selectivity assessment 75
2.3.4 Characterisation of eqMbCF$_3$ 81
2.3.5 Expansion of the substrate scope 83
2.4 Proof-of-principle PrOF experiments 93
2.4.1 Annexin V - ligand state 93
2.4.2 Myoglobin - ligand screening 95
2.4.3 Lysozyme - ligand binding quantification 97
2.5 Perspectives for radiofluorination 98
2.6 Discussion and conclusions 100
2.7 Experimental 105
2.7.1 General information 105
2.7.2 Experimental procedures 113
2.8 Appendix 126

3 Towards the radiosynthesis of ^{18}F-trifluoromethanesulfinate and protein radiolabelling 128

3.1 Introduction 130

3.1.1 [^{18}F]CF_3 versus $CF_3SO_2^-$ synthesis 131
3.2 ^{19}F feasibility studies 135
3.2.1 $BzSCF_2Br$ as a substrate for nucleophilic displacement 135
3.2.2 $:CF_2$ mediated S-S trifluoromethylation 139
3.2.3 Verification of the oxidation-elimination sequence 143
3.3 ^{18}F radiochemistry 146
3.3.1 Radiolabelling of the radical precursor 146
3.3.2 Automation and optimisation of the oxidation-elimination sequence 152
3.3.3 Protein radiolabelling 158
3.4 Conclusions and discussion 161
3.5 Experimental 163
3.5.1 Feasibility studies 163
3.5.2 Radiochemistry 171
3.6 Appendix 177

4 Protein modifications with difluoroethyl radical precursors 182

4.1 Introduction 184
4.2 NaDFES as an N-acetylating agent 185
4.2.1 Initial observations 185
4.2.2 Elucidating the mechanism 187
4.2.3 Potential applications in mild protein acylation 193
4.3 NaDFES as a C-acetylating agent 197

4.4 Alternative •CF_2CH_3 generation system 199
4.4.1 Photoredox activation of difluoroethyl pyridyl sulfone 199
4.4.2 Investigating protein difluoroethylation 203
4.5 Conclusions and discussion 207
4.6 Experimental 212
4.6.1 General information 212
4.6.2 Experimental procedures 214
4.7 Appendix 224

5 C-Formylation of Proteins via Sequential Difluoromethylation Hydrolysis 225

5.1 Introduction 227
5.2 Small molecule model system 229
5.2.1 Establishing LC-MS assay for the AA investigation 229
5.2.2 Verifying reactivity 232
5.3 Peptide model 235
5.3.1 Mechanistic investigations on a model peptide 235
5.3.2 Peptide model for probing selectivity and N-reactivity 240
5.4 Protein chemistry 245
5.4.1 Difluoromethylation 245
5.4.2 Hydrolysis 252
5.4.3 Conjugation of aldehyde tagged proteins 258
5.4.4 Fluorescent properties of aldehyde tagged proteins 266

5.5 Discussion and conclusions 269
5.6 Experimental 273
5.6.1 Experimental procedures 274
5.7 Appendix 284

6 Outlook and perspectives for radical fluoroalkylations of proteins 294

6.1 Perspectives for protein fluoroalkylation 296
6.1.1 Benzothiazole sulfones for radiofluorination 297
6.1.2 Pyridyl sulfones as fluoroalkyl radical precursors 300
6.1.3 Alkyl difluorobromides as fluoroalkyl radical precursors 303
6.2 Conclusions 307
6.3 Experimental 309

References 312

Acknowledgements

My first thanks go to **Prof. Ben Davis** for giving me an incredible opportunity to work with him and to explore the fascinating world of protein chemistry. I am grateful for his unfading support of my scientific endeavours and all that I have pursued during my 4 years in Oxford. Similarly, I would like to acknowledge **Prof. Véronique Gouverneur** for all her kind advice - I still feel lucky to have been able to learn from such inspiring supervisors. None of this would be possible without their engagement in the ISOTOPICS network, a Marie Skłodowska Curie action funded by the European Union, which provided the financial support for my research. It was a great collaborative project that came to life owing to the work of many people, especially: **Dr Christophe Dugave** and **Dr Karen Hinsinger**. Due to this initiative, I was able to realise part of my research in Belgium - I am thankful to **Prof. André Luxen** and **Dr Christian Lemaire** for having me in Liège and training in ^{18}F-radiochemistry. On the other hand, I show my gratitude to the Royal Society of Chemistry, Oriel College and the University of Oxford for their support towards my participation in various conferences. Equally, I am obliged to the Department of Chemistry staff - especially members of the NMR and MS facilities - **James Wickens** for starting my passion for mass spectrometry and **Elisabete Pires**, **Dr John Walsby-Tickle** and **Dr Nader Amin** for all their technical assistance.

Next, I would like highlight the external collaborators: **Dr Osman Tack** and **Dr**

Choon Wee Kee for their interest in radiolabelling of biomolecules. I thank **Prof. Andrew Baldwin**, as well as **Dr Gogulan Karunanithy**, for their time spent in discussing and exercising protein observed ^{19}F NMR experiments.

This leads me to the past and present members of BGD group - there are too many people I would like to thank for their suggestions, help and encouragement. I really appreciate the input of **Dr Patrick Isenegger**, whose unrivalled perspective on science led to many fruitful discussions and profoundly influenced the direction of this thesis. I acknowledge the most interesting flatmate and glycoprotein enthusiast **Dr Ben Bower** for providing a sample of a precious sugar. The ICL squad - for creating an amazing environment to do science - **Adeline Poh**, **Dr Naseer Khan**, **Prof. Charlie Fehl** and others named elsewhere - I will definitely miss that. That does not mean there were not great people in CRL: I salute my favourite post-doc **Dr Jitka Dadova** for sharing many tips, optimism and a protein sample and **Dr Gustaf Hemberg** - the unmatched chess opponent. I would also like to thank everyone in my ISOTOPICS cohort for enduring the Trainings together, and particularly: my fellow labmate **Alex Gafitescu**, my fellow Orielensis **Anna Chiara Vicini** and my dear companions in Belgium **Dr Laura Trump** and **Dr Agostinho Lemos**. Last but not least, I am indebted to everyone who took up the challenge of proof-reading this thesis and particularly **Dr Andrew Giltrap** and **Brian Josephson**. It is also them who taught me great lessons in organic chemistry and biology, respectively.

Thanks to all of my friends back from Poland for their efforts to keep in touch and take breaks across the world. Finally, I could not have done it without **Rubén** and my **parents**. Your love and unconditional support got me through this degree.

Abbreviations

βME	2-mercaptoethanol
A5	annexin V
AA	amino acid
AGC	automatic gain control
Ala	alanine = A
ApoEqMb	apomyoglobin
Arg	arginine = R
Asn	asparagine = N
Asp	aspartic acid = D
ATR	attenuated total reflection
BSA	bovine serum albumin
Bz	benzothiazole
CD	circular dichroism
conc.	concentration
Cys	cysteine = C
DCE	1,2-dichloroethane
DCM	dichloromethane
DDA	data dependent acquisition
Dha	dehydroalanine
DIEA	*N*,*N*-diisopropylethylamine
DMF	dimethylformamide
DMI	1,3-dimethyl-2-imidazolidinone
DTT	1,4-dithiothreitol
EPL	expressed protein ligation
EDT	1,2-ethanedithiol
eqMb	equine myoglobin
ESI	electrospray ionization
EWG	electron withdrawing group
FA	formic acid
FWHM	full width at half maximum
Gln	glutamine = Q
Glu	glutamic acid = E
Gly	glycine = G

HBTU	*N*,*N*,*N*',*N*'-tetramethyl-*O*-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine = H
Ile	isoleucine = I
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
KAHA	α-ketoacid-hydroxylamine amide-forming (Ligation)
LB	lysogeny broth medium
LC-MS	liquid chromatography-mass spectrometry
Leu	leucine = L
Ly	lysozyme
Lys	lysine = K
Met	methionine = M
MS	mass spectrometry
NaDFES	sodium 1,1-difluoroethanesulfinate
NATA	*N*-acetyl-L-tryptophanamide
NaTFMS	trifluoromethyl sulfinate
NCL	native chemical ligation
NDC	non decay corrected
NFSI	*N*-fluorobenzenesulfonimide
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
OD	optical density
PanC	pantothenate synthetase
PBS/Lys	phosphate saline buffer with 100 mM lysine
PET	positron emission tomography
Phe	phenylalanine = F
prec.	precipitation
Pro	proline = P
PrOF	protein observed ^{19}F NMR
PTC	phase transfer catalyst
PTM	post translational modification
RCC	radiochemical conversion
RCY	radiochemical yield
Rel. Exp.	relative exposure

rt	room temperature
RT	retention time
Sar	sarcosine
SAS	solvent accessible surface
sat.	saturated
Ser	serine = S
SET	single electron transfer
SOMO	singly occupied molecular orbital
SPPS	solid phase peptide synthesis
SRP19	signal recognition particle 19 kDa
TCEP	tris(2-carboxyethyl)phosphine
TEMPOL	4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thr	threonine = T
TIC	total ion count
TLC	thin layer chromatography
Tr	porcine trypsin
Tre	trehalose
TRIS	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Trp	tryptophan = W
Trt	trityl
Tyr	tyrosine = Y
UV	ultraviolet
Val	valine = V
XIC	extracted ion chromatogram

Chapter 1

Introduction

Abstract

The chemical modification of proteins has provided a toolbox enabling many experiments aimed at elucidating biological phenomena. In the currently available repertoire, methods targeting native residues are extensively used due to their low operational requirements. However, few strategies are able to specifically target the native (hetero)aromatic residues and rarely directly for the purpose of protein fluorination. Proteins containing fluorine are gaining popularity due to their advantageous applications: such as in protein observed ^{19}F NMR, non-invasive PET imaging or physicochemical properties engineering. Currently, the use of this tag is limited by the scarcity of the methods for its installation and reliance on biosynthetic protocols. The development of new methods in organic synthesis, suggested the possibility for fluoroalkyl radical utilisation for the direct modification of proteins. This chapter investigates the above topics and lays the foundation for discussion on how and why fluoroalkyl radical chemistry could be applied in protein chemistry.

1.1 Chemical modification of proteins

Proteins are at the core of the structural and regulatory machinery of every living system, being indispensable for biochemical processes. Although limited to a rather small set of canonical amino acid building blocks, these biomolecules can perform various functions depending on combinatorial linear arrangements that allow them to adopt diverse spatial conformations[1]. Furthermore, Nature has found additional benefits in introducing chemically diverse post-translational modifications (PTMs) to their residues. This expanded chemical diversity allows more precise control of protein roles in biological processes and circumvents the limitation of the restricted genome[2].

Therefore, it does not come as a surprise that, taking this general inspiration from Nature, chemists have been investigating methods for the chemical modification of proteins for decades, whether driven by the desire to better understand the structure-activity relationship[3,4], or to alter their chemical and physical properties and hence the native function[5,6]. To this end, various strategies have been explored - Fig. 1.1.

- Recreating native PTMs purely chemically has been shown to be a considerably difficult task[7], typically achieved by segments assembly. It is aided by methods such as Native Chemical Ligation (NCL) or Expressed Protein Ligation (EPL), which allow for the construction of physiologically relevant modified proteins e.g. phosphorylated[8]. Although requiring considerable effort, such an approach enables, in principle, the creation of the biomolecules from scratch e.g. total synthesis of homogeneous glycopeptides[9].

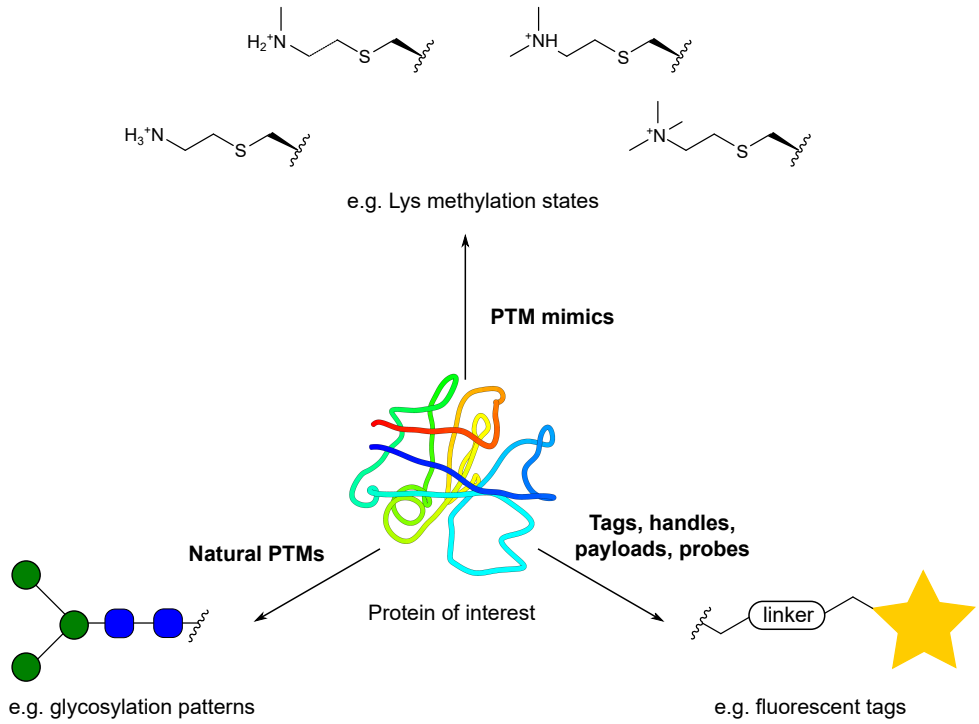


Fig. 1.1: Examples of protein chemical modifications that can be useful for studying biological systems.

- The introduction of PTM mimics, often with unnatural linkers/replacements, is nonetheless useful due to the chemical ease with which they are installed. Importantly, this has allowed incorporation of functional structural mimics e.g. lysine (Lys) methylation states[10] (using formal C $\rightarrow$ S replacement) or glycosylation patterns[11], becoming an important field greatly contributing to the expansion of our knowledge about the role of specific PTMs in living organisms.

- Finally, various methods have been designed to allow protein 'labelling' i.e. incorporation of often unnatural probes[12] (of activity, localisation etc.), tags[13] (for identification, affinity etc.), reactivity handles[14] (for immobilisation etc.), payloads (for therapeutic/diagnostic applications)[15] and other functionalities[16],

useful in protein studies.

Many solutions to the challenge of creating recombinant proteins have been provided, opening avenues for chemical biology for studying fundamental biological problems. This introduction will briefly discuss examples of available tools for the in vitro post-translational chemical protein modification with a particular focus on the modification of native residues, methodologies using radical-based chemistry and the general topic of installation of fluorinated residues into proteins.

1.1.1 Modification strategies for engineered proteins

A substantial part of current biological/biomedical research uses recombinant proteins, which are produced via in vitro expression systems (e.g. bacterial) and purified even up to mg-like quantities[17]. Such an approach allows a certain degree of pre-engineering of the target protein that may be used to expand choices of possible modification strategies. The biomolecule can be redesigned with a unique (un)natural insert of a specific reactivity profile that can be placed at a desired site. Dubbed 'Tag-and-Modify', this strategy can produce excellent results at the expense of the tag introduction[18]. This can be achieved with a multitude of biosynthetic methodologies with varying degrees of labour including:

- site-directed mutagenesis[19],
- global replacement of a residue analogue in an auxotrophic expression host e.g. methionine[20] (Met) or phenylalanine[21] (Phe),
- genetic code expansion (e.g. amber codon suppression/reassignment)[22],

- enzymatic incorporation[13] and many other strategies.

The last decade saw an explosion in the number of new methodologies that rely on this strategy. Most of them stem from the adaptation of established organic chemistry reactions and their translation to biocompatible conditions (buffered aqueous solutions at ambient temperature, neutral pH), provided that they offer bioorthogonal reactivity i.e. there is no cross-reactivity with functional groups natively present in proteins (amines, carboxylic acids, alcohols etc.) and favourable kinetics (as the typical biomolecule concentration in a reaction mixture lies in the low $\sim \mu$M range)[23]. It is also essential that the employed conditions do not compromise protein integrity, allowing the preservation of the native structure and biological function. Along with the development of biosynthetic methodologies, the chemical space of possible tags structures is ever growing, allowing customised chemistry that can be utilised for a given specific experiment. General trends were excellently summarised in recent reviews by Bernardes *et al.*[24,25]. A brief outline of the commonly employed tags and the subsequent chemistry used for their modification is given below and in Fig. 1.2.

This list is by no means exhaustive and serves to provide a general outlook of the chemical toolbox available for chemical protein modification. It is worth noting that these strategies are usually designed to complement the reactivity offered by native residues and to enable bioorthogonal chemistry even in complex protein mixtures or in vivo applications.

1. It is often convenient to mutate a specific amino acid in a protein of interest to one that is otherwise absent or not solvent accessible and use its native reactivity

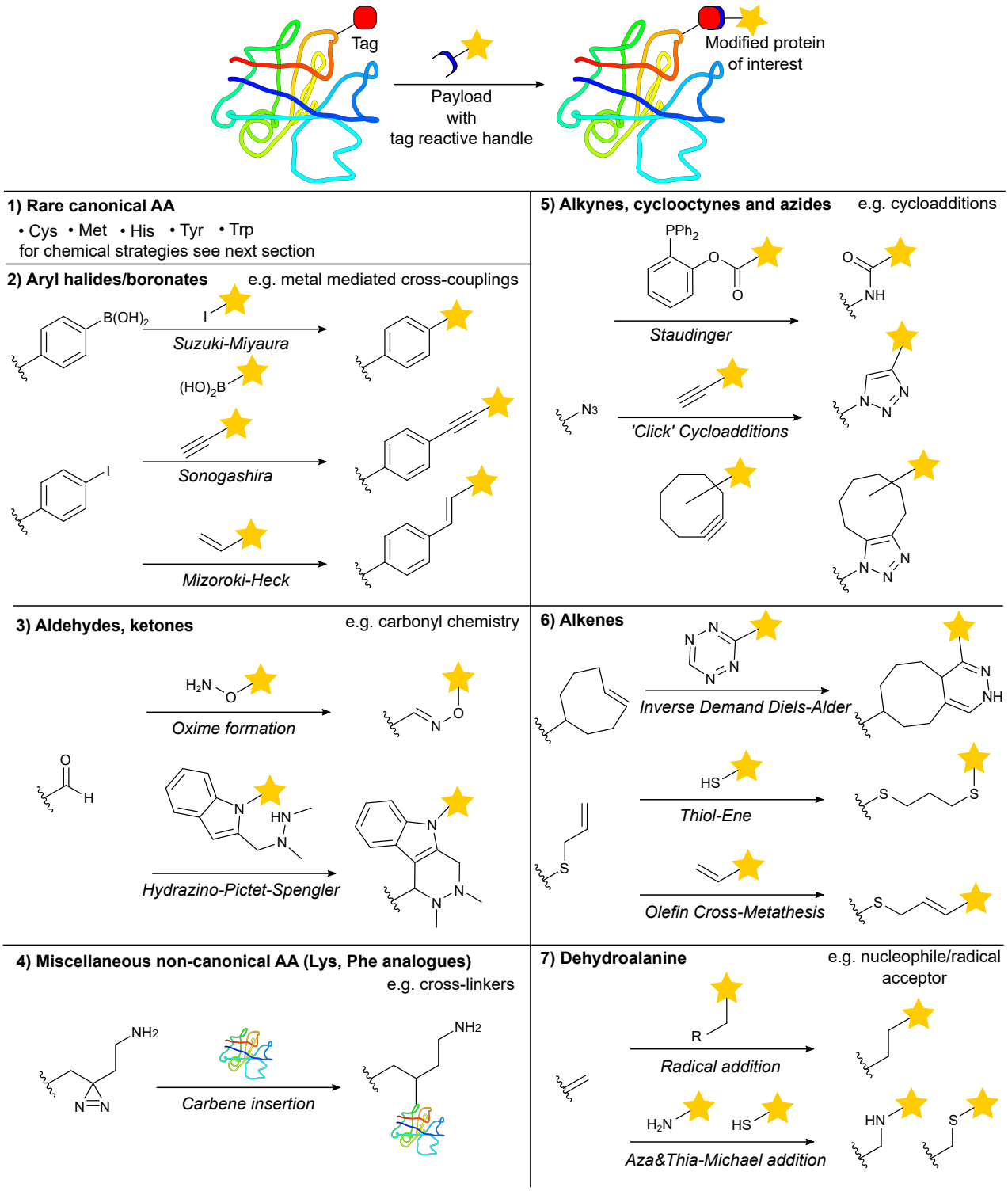


Fig. 1.2: Selected examples of handles used in the 'Tag and Modify' strategy for engineered protein modifications[24,25].

for introducing a modification. This can be performed for a range of residues, but is most straightforward and often employed for cysteine (Cys) labelling e.g. for probing enzymatic activities upon ^{19}F tagging[26,27]. Typically in folded proteins these residues are not only rare but also can be engaged in disulphide bonds (especially for non-cytosolic proteins)[28] and hence are unreactive in free thiol specific chemistry. Another advantage is the versatile Cys reactivity that will be discussed in a greater detail in the next section.

2. Among tags that can be introduced via biosynthetic methods, aryl halides and boronates were found to be suitable substrates for metal mediated cross-couplings[29], which allowed e.g. the in-depth study of the glycogen formation process[30]. Key to the success of this chemistry was a new class of Pd ligands that made it possible to achieve good reactivity in aqueous media[31].

3. Valued for their versatile bioorthogonal reactivity, small size and relative large number of installation methods, aldehydes and ketones are well explored tags for protein modification[32]. Notwithstanding, their extreme utility in protein labelling still provokes research and discovery of new protein compatible chemistries, e.g. to get around the limitation of reversibility of formed constructs (Pictet-Spengler reaction)[33]. Through the reactivity towards α-nucleophiles and other carbonyl group specific chemistry, affinity/fluorescence labelling was realised which enabled e.g. powerful single molecule studies[34,35].

4. The technology of genetic code expansion allows for incorporation of a variety of structural analogues of certain amino acids, best realised for (pyrro)lysine and

tyrosine residues[36]. Although this method requires a substantial amount of the synthetic small molecule to be incorporated and is often limited in efficiency, it achieves remarkable selectivity and, importantly, it can be applied to in vivo systems. Prominent applications include the introduction of photo reactive groups such as diazirine rings that upon light irradiation promote proximity dependent cross-linking with interacting protein partners in native settings[37].

5. Because of their fast kinetics, orthogonality and high yields, azide–alkyne cycloadditions have been the centrepiece of the field named 'click chemistry'. Possessing such characteristics, there are numerous examples of their applications in chemical biology[38], enabling e.g. labelling of proteins with short-lived ^{18}F isotope for imaging purposes[39]. This methodology has been refined to eliminate the need of toxic Cu(I) as a catalyst, achieved by changing the coupling partner to a strained alkyne[13]. This greatly facilitated in vivo applications, not only for protein labelling[40], but also glycans modification[41] or selective tagging of cancer cells[42]. Azides can also be utilised in the Staudinger ligation[43], notably in its traceless version[44] further highlighting the value of this 1,3 dipole tag in the field of bioorthogonal chemistry. The protocol can also be easily reversed by the incorporation of an alkyne tag into the protein.

6. Another important class of tags are alkenes. Various (strained) alkenes (dienophiles) turned out to be applicable in [4+2] cycloadditions typically with tetrazines (dienes) in a class of inverse-demand Diels–Alder reactions[45]. Notably, this reactivity is characterised by fast kinetics and often leads to activation of fluo-

rogenic probes and therefore has found many applications throughout chemical biology e.g. a chemical-caging/rescue of active residues[46]. On the other hand, alkenes in a side chain were successfully demonstrated e.g. to undergo olefin cross metathesis for protein PTM mimic writing/erasing[47] or thiyl radical addition for ubiquitination[48].

7. Easily accessible from Cys, as well as other residues, dehydroalanine (Dha) has been shown to be a versatile tag offering the possibility for linker size reduction[49]. Its specific stereoelectronic properties makes it an excellent acceptor of both N and S nucleophiles (among others). The Aza-Michael reactivity has been optimised to a level where a stoichiometric level of substrate can be used, which is still rare in protein chemistry[50]. The mechanism of the 1,4-additions was studied in-depth and was exploited to realise protein deuteration[51]. Another prominent mode of Dha reactivity is its SOMO-phile character and subsequent captodative stabilisation of the formed α radical permitting ready modification with C-centred radicals[52]. This was showcased by the development of a broad substrate scope post-translation chemical mutagenesis methodology[53].

Although many methodologies are already available, it is key that they should be operationally accessible/viable to the wider community of researchers. Only then they can truly make a contribution for discovery of fundamental biology, imaging and therapeutic applications.

1.1.2 Strategies for native residue modification

From this picture it is evident that the extra layer of complexity, provided by the installation of a chemical handle through genetic manipulation, can ensure both superior selectivity and reactivity. However, this strategy is not viable in all cases as sometimes the operational complexity, costs or protein sourcing preclude its use. On the other hand, the reactivity provided by the native residues may be sufficient for the purposes of the experiment, rendering more complex approaches uneconomical.

For this reason, the reactivity offered by native residues has been the object of particular attention. Methods based on it are by definition more straightforward and easily applicable, but often limited in versatility and product homogeneity. Nevertheless, these advantages have allowed for the biomolecular constructs generated in this way to be advanced for clinical applications[54]. Such approaches are also attractive since they can usually be easily translated between different protein systems.

Discussed previously, the simple adaptation of established organic chemistry methods has not been possible to achieve modification of folded proteins and, if only, is applicable to simpler peptide sequences where the number of cross-reactive functionalities is lower and use of protecting groups possible. Even in that case, the repertoire of residues that can be efficiently targeted in a chemo- and site-selective manner is somehow limited[55]. A general overview of the available strategies (both established and emerging, however not comprehensive) is presented in Fig. 1.3 and discussed below.

- There are scarce reports of methods targeting peptidic aliphatic (alanine (Ala),

glycine (Gly), proline (Pro), valine (Val), leucine (Leu), isoleucine (Ile)) with an example of Pd-mediated C-H activation of Ala side chain[56], while on proteins, residual modification is observed during radical footprinting[57].

- Glutamine (Gln) and asparagine (Asn) are also among other residues that are not typically targeted and their modification requires special proximity driven reactivity strategies e.g. exploiting a rhodium metallopeptide catalyst[58].

- Similarly, methods for serine (Ser) and threonine (Thr) suffer from the chemoselectivity issues (as in simple esterifications[59]). However, selective modification of Ser was realised with similar carbenoid based strategy[60], or only very recently using oxathiophospholane reagents, regrettably suffering from the requirement of the organic solvent use[61].

- Although aspartic acid (Asp) and glutamic acid (Glu) possess a well defined reactive handle (e.g. for amide bond formation[62]) their relatively high abundance (with the additional presence of the C-terminal carboxylate) and cross-reactivity make them difficult targets for specific labelling[63]. Selective modification can be partially enabled by the distinct structure of the target protein[64].

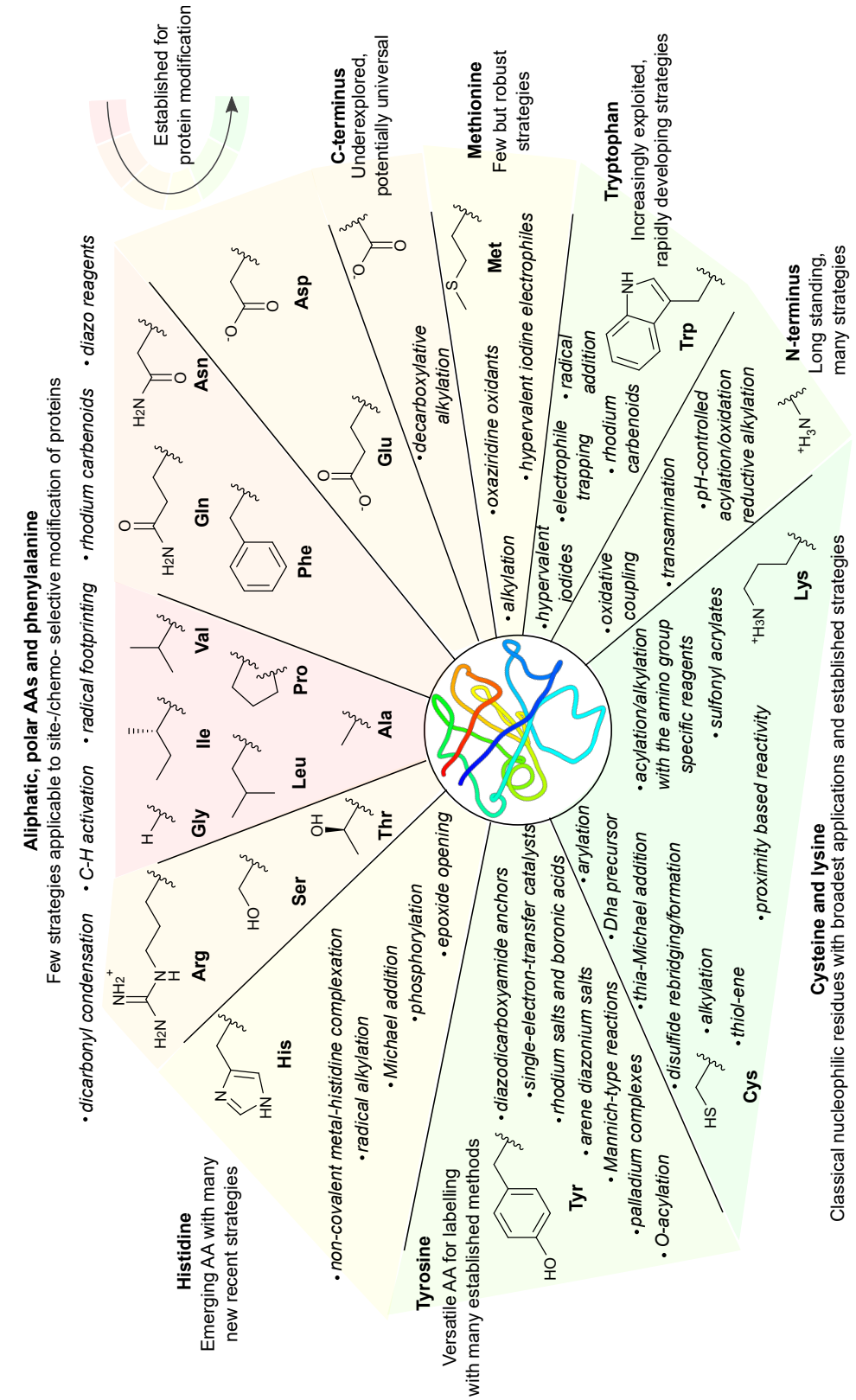


Fig. 1.3: General overview of strategies for the modification of native residues[65].

- Due to the relatively low nucleophilicity of arginine's (Arg) side chain guanidinium group ($pK_a > 12$), competition with native amines and rather high abundance, these residues have not been extensively used for protein modification, with an exception of glyoxal-based strategies[66].

- Lastly, because it possesses the lowest reactivity among the aromatic residues, chemoselective phenylalanine (Phe) modification strategies for proteins are nearly absent.

Similarly rare is the modification of the backbone chain that was only possible using the directing ability of a neighbouring histidine (His) residue[67]. Modification strategies for commonly targeted residues are reviewed in a greater detail below, with a particular focus on emerging strategies.

Histidine

One of the AAs that has been traditionally difficult to target in a selective manner but has seen a lot progress in recent years is histidine. The non-covalent complexation via extended His-tag motifs, extensively utilised in the purification of proteins[68], has also been used for selective protein modification with fluorophores[69] - Fig. 1.4a.

The reactivity of the imidazole ring as a nucleophile was initially exploited for an epoxide opening - Fig. 1.4b - in an interesting site-specific proximity guided manner[70], and also with an imine replaceable spacer[71]. It is worth noting that these strategies (like most for modification of native residues) are limited by the solvent accessibility of His residues to be labelled, which sometimes can be advantageous in achieving surface site-specificity[70].

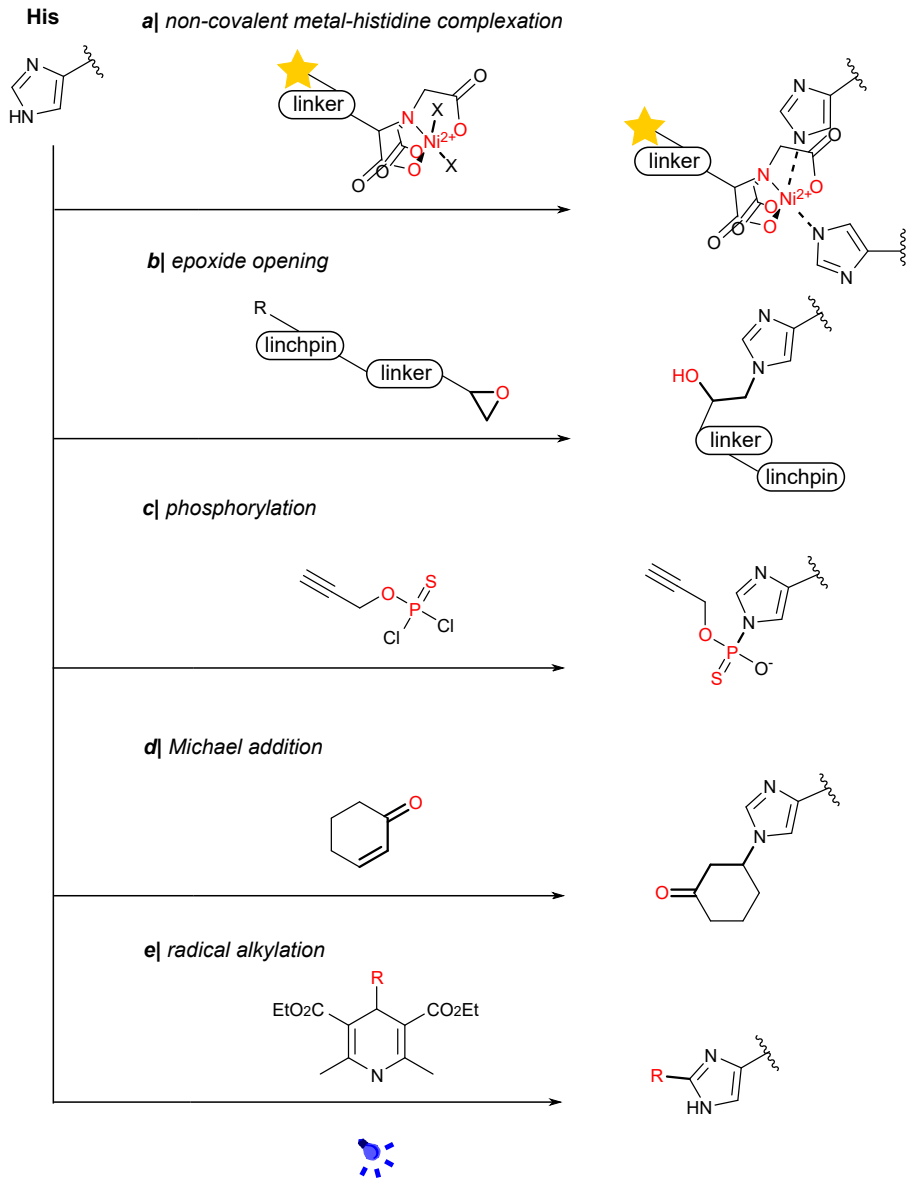


Fig. 1.4: Overview of the strategies for chemoselective His modification: **a**| Ni^{2+} tris-nitrilotriacetic acids for non-covalent complexation[69], **b**| epoxide as a His-specific electrophile in linchpin reagents (with a protein specific ligand[70]/aldehyde tag[72]), **c**| thiophosphorodichloridates for His-specific phosphorylation[73], **d**| 2-cyclohexenone as a Michael acceptor[74], **e**| light activated *C*4-alkyl-1,4-dihydropyridines as radical precursors for Minisci chemistry[75].

In the absence of a ligand for the protein of interest, the linchpin strategy can be used, where His-epoxide reactivity is proximity enhanced by attachment of an aldehydic

handle to the reagent allowing reversible Lys binding[72].

In recent years a more diverse set of His selective electrophiles has been discovered. Thiophosphorodichloridate reagents allow ready phosphorylations[73] - Fig. 1.4c and 2-cyclohexenone effective Michael additions[74] - Fig. 1.4d. Alternatively, site-selectivity is achieved via bis-alkylation of a short His-tag sequence[76]. On the other hand, an example of imidazole selective metal complexations (with platinum) was also reported[77]. These examples demonstrate considerable progress, but offer only limited conversions, likely because of the fine balance between reactivity/selectivity, struck to avoid competing reactions with other nucleophilic residues (Cys, Lys).

Another notable strategy, leading to the formation of a C-C bond rather than C-N bond, is the radical alkylation of His residues using Minisci chemistry[75] - Fig. 1.4e. This method has started being adapted for biomolecules, however, so far has only been successfully applied to peptides and certain proteins due to the need for acidic conditions and problems with the mild radical generation in aqueous solution[78].

Tyrosine

Tyrosine is another AA that, classically, has not been successfully modified in a chemoselective manner. However, rapid progress in the last two decades has made it an established target for protein modification[79].

Several strategies based on derivatisation of the phenoxide anion have been developed e.g. acylation, however, the residue selectivity of this chemistry is typically low, with best results claimed to be achieved with reagents like *N*-acetylimidazole[80]. The technology was further developed with better acyl transfer reagents, and also modified

with a ligand to achieve the site specificity[81] (4-dimethylaminopyridine catalysed acyl transfer from thiophenyl esters) - Fig. 1.5a. O-modification (specifically alkylation) could also be realised via the use of π-allylpalladium complexes[82].

Similar ligand-tethered strategies can also be based on different chemistry. To benefit from the specific electronic properties of Tyr, a single electron transfer catalyst $[Ru(bpy)_3]^{2+}$ (ruthenium tris(2,2'-bipyridyl) complex) was attached to a ligand, which upon light irradiation promoted the formation of the radical on proximal Tyr residues that can be subsequently modified with specific trapping agents[83] - Fig. 1.5b. However, it was found that this reactivity is not limited to Tyr residues as it also results in Cys labelling[84]. This type of reactivity resembles that of oxidative couplings where in the presence of a strong metal oxidant, non-specific O- and *ortho* C-coupling to electron rich anilines was possible[85].

Another notable reactivity of Tyr that has been adapted into protein chemistry was the ene-type reaction with diazodicarboxylates[86]. These reagents, although suffering from reactivity/stability issues under aqueous conditions, were optimised for protein labelling in a form of 4-phenyl-3*H*-1,2,4-triazole-3,5(4*H*)-dione scaffold - Fig. 1.5c. Improvement to this technology was brought by the use of electrochemical activation of the precursor reagent, which significantly increased observed conversions[87].

Aryl diazonium salts belong to another class of compounds with a known affinity towards Tyr - Fig. 1.5d. These modifications have varied, but predominant, selectivity towards Tyr that depends on the substitution pattern in the aromatic system. Particularly good results were achieved with a carbonyl group, either as a ketone[64] or

an aldehyde[88] with an additional advantage being their possible use for a subsequent conjugation.

Successful modification of Tyr was achieved through three-component reactions using Mannich type chemistry - Fig. 1.5e. An imine, formed in situ from an aldehyde and an electron rich aniline, could be trapped by native Tyr residues in proteins[89], although participation of Trp side chains was also observed[90].

In another three-component reaction, a Rh^{3+} salt was used to conjugate a protein with a range of aromatic groups from aryl boronic acids precursors - Fig. 1.5f. High conversions were only possible with the privileged *ortho* carboxamide scaffold and the modification was reversible in the presence of 1,4-dithiothreitol (DTT)[91].

In a recent interesting report, Tyr was used as an acceptor for an electrogenerated phenothiazine N-radical leading to an *ortho* substituted arylamino product[92] - Fig. 1.5g.

Taking advantage of the natural biochemical reactivity of Tyr, an enzymatic method was developed in which the tandem use of two enzymes allowed methoxylation of accessible Tyr residues, however further advances are needed in order to introduce more diverse tags on a broader substrate scope[93].

It is worth mentioning that a popular strategy is a multi-step chemical modification, in which a reactive handle is introduced in the first step, which gives access to the different bioorthogonal chemistries. Examples of these strategies include the formation of an aza-coupling product with diazonium salt and subsequent reduction to an *ortho* aminophenol derivative followed by specific for these compounds chemistry like oxidative couplings[94] or iodination for purposes of Pd-catalysed cross couplings[95].

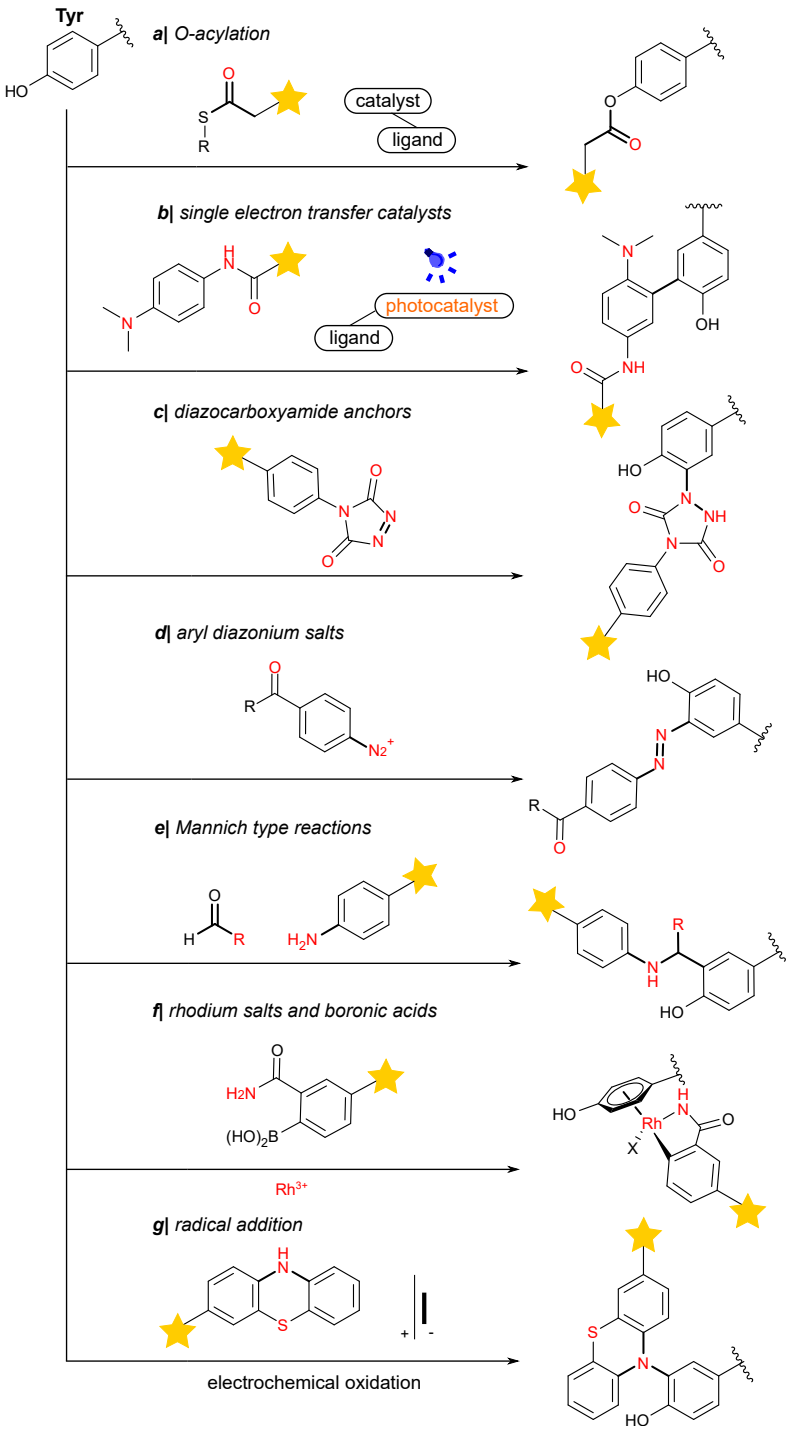


Fig. 1.5: Overview of strategies for chemoselective Tyr modification: **a**| thioesters as acyl donors for O-acylation directed through a ligand/nucleophilic catalyst system[81], **b**| ligand/photocatalyst system for a proximity driven formation of tyrosyl radicals readily trapped with electron rich anilines[83], **c**| diazodicarboxylates as Tyr specific ene-partners[86], **d**| carbonyl benzodiazonium salts as potent electrophilies for Tyr modification[64,88], **e**| Mannich reaction between an aldehyde, electron rich aniline and Tyr[89,90], **f**| rhodium salt promoted aryl boronic acid conjugation strategy[91], **g**| phenothiazine derived radical addition[92].

This overview of the available strategies highlights the high versatility of Tyr as the target residue for protein modification. Some restrictions stem from its partial amphiphilic character that results in a diverse distribution between the solvent accessible surface and the blocked interior of the protein, as well as from the selectivity issues due to the competition from other aromatic residues, particularly Trp.

Methionine

Met has been an obvious target for protein modification since it is the only residue with substantial nucleophilic character at acidic pH[96]. As such, various alkylating agents could be employed but one of the best stability/reactivity patterns was found for reagents giving benzylic sulfonium products, that could also be easily dealkylated with 2-mercaptopyridine[97] - Fig. 1.6a. However, extremely low pH (< 3) was needed to achieve Met selectivity, which makes this method less relevant for protein modification.

Similarly, useful protein–sulfonium conjugates could be formed with hypervalent iodine(III) reagents with improved kinetics - Fig. 1.6b. Particularly effective constructs were achieved with an optimised reagent bearing ethyl diazoacetate motif that not only allows the introduction of a functional payload but also provides an electrophilic diazo probe[98]. Again, the need for acidic conditions limits the general applicability.

Significant improvement was achieved with the introduction of redox based reagents allowing chemoselective strain-driven Met sulfur imidation at neutral pH - Fig. 1.6c. Oxaziridines with a weakly electron withdrawing urea turned out to be key to sup-

press the undesired O-transfer product and allowed N-conjugation with an alkyne handle[99].

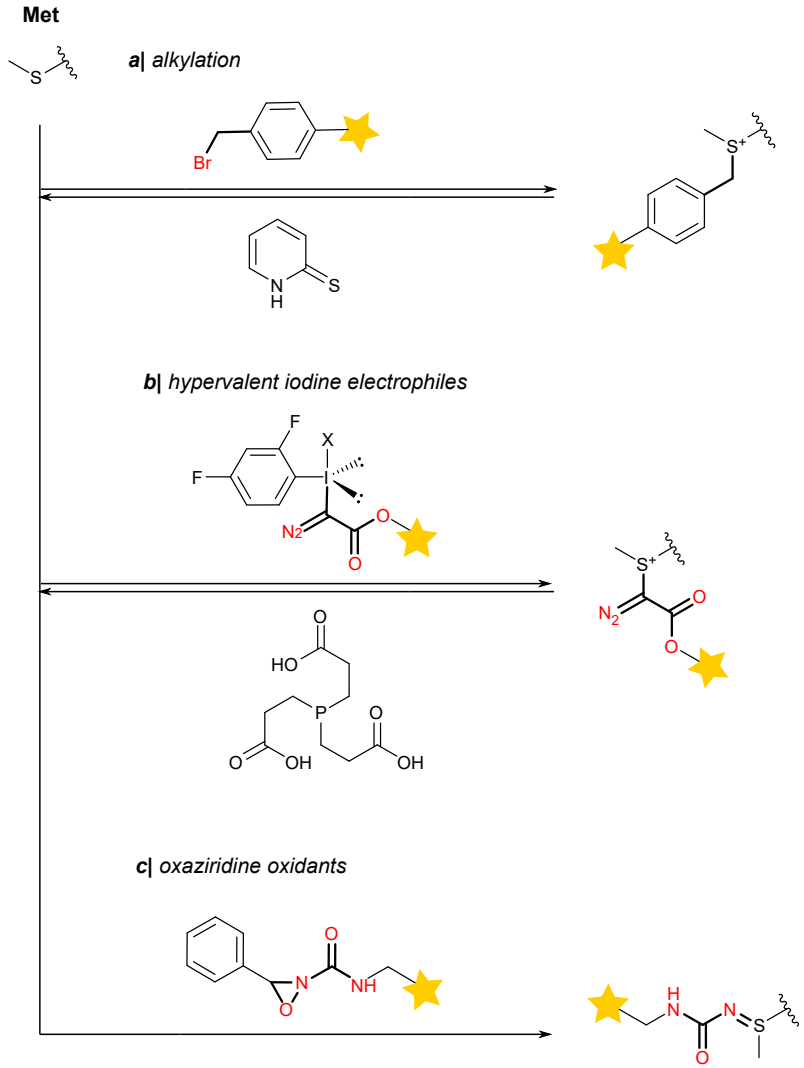


Fig. 1.6: Overview of selected strategies for chemoselective Met modification: **a**| benzylic alkylating agents[97], **b**| electrophilic hypervalent iodine(III) reagents with ethyl diazoacetate motif[98], **c**| oxaziridines with a functionalised electron withdrawing urea[99].

Although Met oxidation (e.g. to sulfoxide) has been an established practice, currently there are no more protein bioconjugation methods that are based on such chemistry[100].

C-terminus

Although being a relatively obvious target, the protein C-terminus is extremely hard to modify chemoselectively over other carboxylates present in a protein. A significant breakthrough was reported recently in which the oxidation potential was used as the discriminating factor[101]. The increased stability of the C-terminal C(α) radical allowed modification with Michael acceptors - Fig. 1.7.

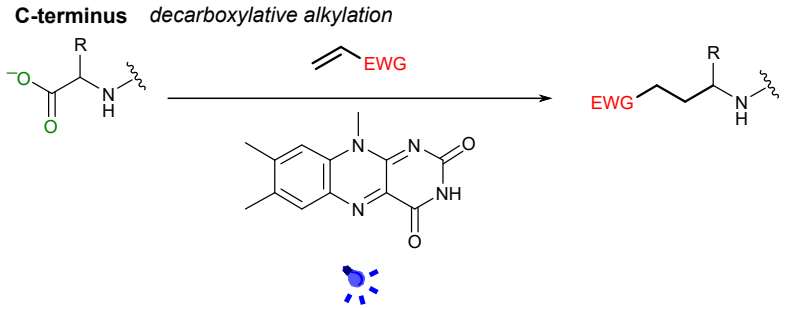


Fig. 1.7: Chemoselective strategy for chemical C-terminal modification: photoredox catalysed decarboxylative alkylation with a Michael acceptor (EWG = electron withdrawing group)[101].

The lack of chemical methods contrasts the number of biosynthetic methods based on enzymatic approaches[102], often requiring the presence of specific sequences.

N-terminus

The N-terminus of proteins is unique as it possesses a generally solvent accessible amine group with distinct p$K_a \sim 8$, sufficiently different from that of Lys residues (p$K_a \sim 10$)[103]. Therefore, some amine specific chemistries (e.g. reactivity with hard electrophiles) can be applied in a chemoselective manner, often just by controlling the pH of the reaction mixture. In this way, the reductive alkylation with various aldehyde partners could be realised - Fig. 1.8a, being universal for all AA side chains

apart from Cys (which reacts to form a thiazolidine - which is a valid modification strategy in itself)[104]. It is a common trend that methods for N-terminal modification are generally tolerant of most AA side chains apart from few. Similarly, it was possible to achieve acylation[105], with an additional layer of specificity added by using a ketene reagent - Fig. 1.8b.

Certain aldehydes are known for their transamination reactivity leading to formation of the N-terminal ketone as a useful chemical handle. Initially, it was found that pyridoxal-5-phosphate was efficient in this transformation[106], but the technology was further improved with the use of the easier to handle Rapoport's salt (*N*-methylpyridinium-4-carboxaldehyde benzenesulfonate salt)[107] - Fig. 1.8c. On the other hand, some other aldehydes were found to give condensation products in high yields via cyclic imidazolidinone formation - Fig. 1.8d.

Imines formed with 2-pyridinecarboxaldehydes were particularly suitable for the nucleophilic attack of the neighboring amide nitrogen necessary for the formation of the product[108]. Easier access to the functionalised reagents was provided by the discovery of the equally reactive 1-(4-nitrophenyl)-1H-1,2,3-triazole-4-carbaldehyde scaffold[109].

In opposition to these general strategies, methods targeting only specific N-terminal AAs have been explored. This has been thoroughly exploited for terminal Thr and Ser[110] residues, as periodate oxidation leads to formation of a carbonyl handle. One of the most important methods for the protein (semi)synthesis - native chemical ligation - relies on the reaction of an N-terminal Cys with a thioester partner[111]. In another example, a method for the specific labelling of a terminal Gly residue via

aminoalcohol formation was developed[112].

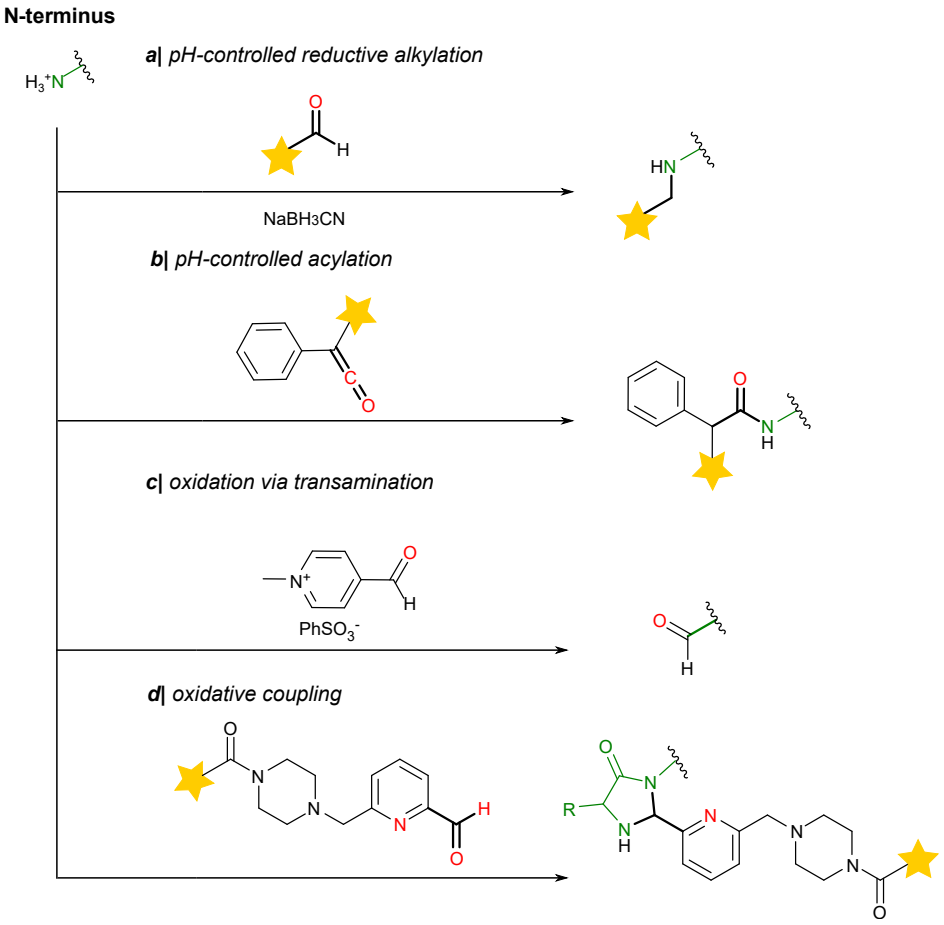


Fig. 1.8: Overview of strategies for chemoselective N-terminal modification: **a**| pH controlled reductive alkylation[104], **b**| ketene donor acylation[105], **c**| transamination with Rapoport's salt for aldehyde handle formation[107], **d**| 2-pyridinecarboxaldehydes for formation of a stable condensation product[108].

Even though the N-terminus would seem to be a universal target, it is important to remember that sometimes it is the site of endogenous modification, and therefore often inaccessible for chemists[113].

Tryptophan

Trp is one of the most useful residues to target because it has the lowest relative abundance ($\sim$ 1%) in proteins[114]. However, many classical methods lack chemose-

lectivity, and some do not lead to constructs useful for labelling (e.g. oxidation)[115]. Others employ harsh conditions in order to achieve high selectivity (e.g. malondialdehyde modification[116]). Among them, the use of rhodium carbenoid intermediates[117] was possible under mild enough conditions to be successfully applied to proteins - Fig. 1.9a.

The methodology was further developed to avoid the need for low pH by changing the additive to a different hydroxylamine[118]. In another example, site selectivity was achieved by preparing a rhodium catalyst metallopeptide allowing for proximity driven reactivity[119]. In every case, the solvent accessibility of the Trp residue to be modified was the main factor affecting the observed conversions.

Another metal mediated strategy was based on the gold catalysed addition of hypervalent iodine reagents - Fig. 1.9b. Using the Waser reagent, introduction of an ethynyl handle was possible, however, the reaction required an addition of organic co-solvent to proceed[120].

On a side note, a photocatalytic approach was used to generate a radical on the β carbon on Trp side chain (skatolyl radical) that could be trapped with a Michael acceptor[121]. However, this method resulted in side reactions at the C-terminus, as described above, and was only demonstrated for peptides.

A more general strategy based on a radical mediated transformation was use of the 9-azabicyclo[3.3.1]nonane-3-one-*N*-oxyl persistent radical reagents as oxidasible precursors allowing C3 modification of the Trp residue[122] - Fig. 1.9c. Although metal-free and rapid, this method requires low pH in order to generate NO_x from the nitrites, which limited it general applicability.

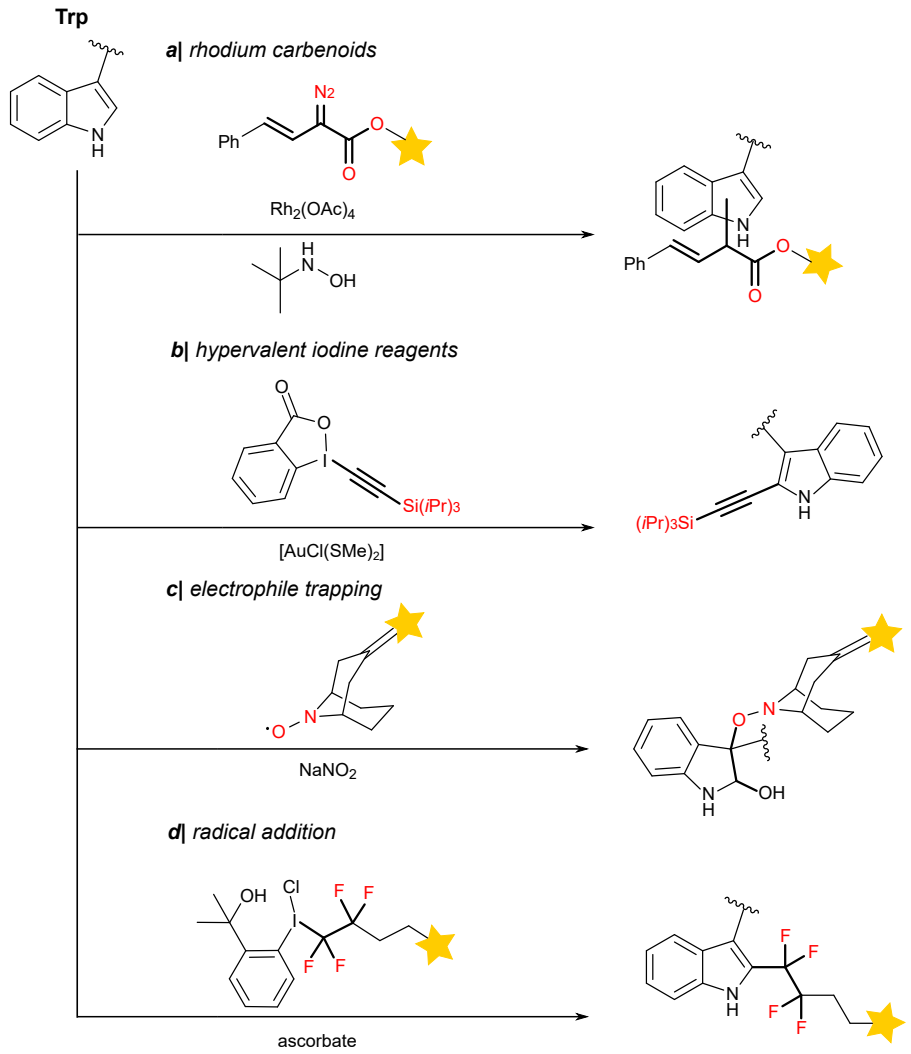


Fig. 1.9: Overview of strategies for chemoselective Trp modification: **a**| rhodium carbenoid intermediates for accessible Trp residues labelling[117,118], **b**| gold catalysed addition of iodanes[120], **c**| organoradical additions to C3 position[122], **d**| hypervalent iodine precursors for radical fluoroalkylations[123].

Finally, the potential of radical chemistry for Trp selective modification was realised with certain fluoroalkyl radicals derived from hypervalent iodine precursors[123] - Fig. 1.9d. Such methodologies will be discussed in a greater detail in next Chapters.

In a recent brief report, heterogeneous biohybrid palladium nanoparticles were used for C2 arylation of a protein Trp with diazonium salt partner, however the

scope in terms of both substrates was not tested[124].

Cysteine

Due to its potent reactivity and relatively low abundance, Cys is one of the most frequently targeted residues for protein modification. Classic methods rely on the ready nucleophilic reactivity of deprotonated Cys residues, with a prominent example of modification with maleimides. However, the conjugates formed in this reaction are known to suffer from poor stability, as it is readily reversible. This has been improved with a new design of maleimide scaffolds[125]. Moreover, the reversibility was used as an advantage in a system where retro-Michael reaction with a 5-methylene pyrrolone substrate could be reversed in a controlled manner through use of a basic buffer (pH > 9) or via thiol exchange at neutral pH[126] - Fig. 1.10a. On the other hand, thiol alkylation could be realised in near stoichiometric ratios and in an irreversible manner using carbonylacrylic reagents[127] - Fig. 1.10b. Other widely used electrophiles include α-halocarbonyls and vinyl sulfones[128].

Site-specificity in a case where multiple Cys residues are present was achieved in a system where the immediate environment (4 AA sequence) around the selected Cys residue was fine-tuned to maximise the interaction with the reagent (by formation of a π-clamp)[129] - Fig. 1.10c. Many strategies have been developed based on this elevated reactivity of the residue within a specific sequence[130,131].

A different type of reactivity was exploited in the so-called thiol-ene and thiol-yne couplings, where the propensity of a sulfhydryl group to form a thiyl radical is capitalised on - Fig. 1.10d.

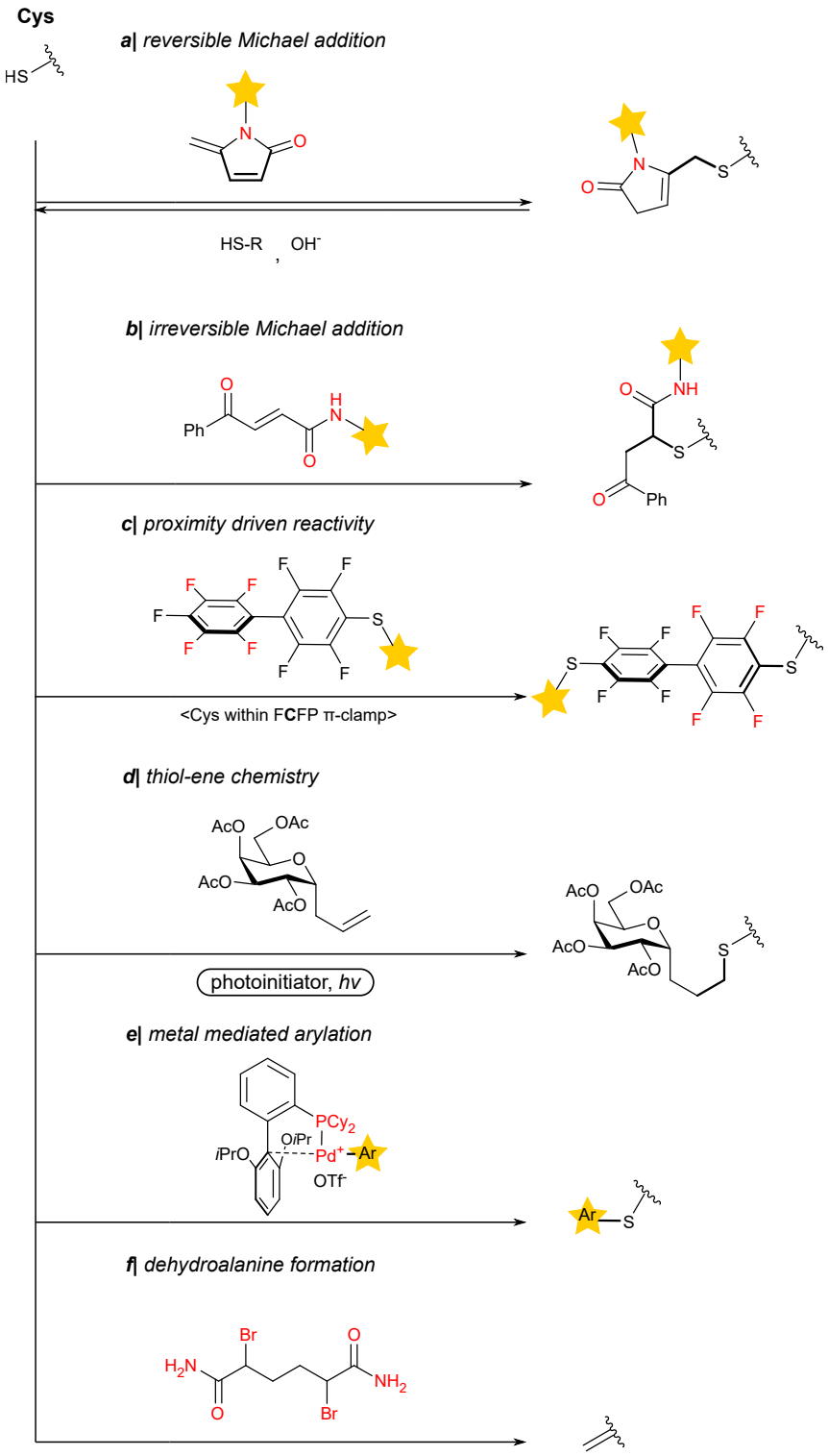


Fig. 1.10: Overview of strategies for chemoselective Cys modification: **a**| 5-methylene pyrrolone reversible alkylation [126], **b**| carbonylacrylic reagents for irreversible Michael additions [127], **c**| site-selective Cys modification with specificity inducing sequences [129], **d**| alkenes as acceptors for protein thiyl radicals [132], **e**| metal catalysed Cys arylation [133], **f**| dehydroalanine formation via bis-alkylation/elimination [49].

Such species (often generated with a photoinitiator) can add to an alkene as well as alkyne (usually leading to double modification), which has been used for protein glycosylation[132] or ubiquitination[134].

Recently, metal catalysis was applied to Cys arylation in a more efficient manner[135]. Taking advantage of the soft nucleophilicity of sulfur in the interaction with a preformed Pd(II) complex, S-aryl conjugates could be obtained under mild conditions[133] - Fig. 1.10e. This technology was applied to achieve intermolecular Cys-Lys biomolecule cross-linking[136]. Similarly, the thiophilic nature of gold allowed ready Cys arylation with a range of Au(III) complexes with faster kinetics as compared to the Pd based strategy[137]. On the other hand, Pd catalysis has enabled Tsuji-Trost allylation of Cys residues, thus allowing the formation of the prenylation PTM mimics[138].

It is worth mentioning that Cys is one of the best precursors to the versatile reactive handle - dehydroalanine[49]. For example, a bis-alkylation/elimination sequence with 2,5-dibromohexanediamide leads to the formation of Dha residues from accessible Cys residues under mild conditions - Fig. 1.10f.

Other strategies rely on disulfide rebridging in which a disulfide bond engaging two Cys residues gets broken and reformed with a cyclic handle for bioconjugation. This method is particularly popular for the modification of antibodies[139].

Lysine

Another residue widely employed in protein chemistry due to its ready nucleophilic reactivity is Lys. Considering its relatively high abundance, site-selectivity has been the

biggest issue. Nevertheless, labelling with amine-reactive reagents (activated esters, iso(thio)cyanates, aldehydes, anhydrides, etc - Fig. 1.11a)[140] has been an established and widespread practice for the attachment of fluorescent tags as well as antibody drug conjugate production[141]. Similarly, most of the previously described chemistry for the N-terminal modification (reductive alkylation, activated esters etc.) can be applied to the ε-NH_2 group of Lys residues. Strategies to achieve site selectivity employ either kinetic control to target the 'most reactive' Lys (due to the modulation of the amino group's pK_a dependent on the local environment)[142] with a notable performance of the sulfonyl acrylate reagents[143] - Fig. 1.11b.

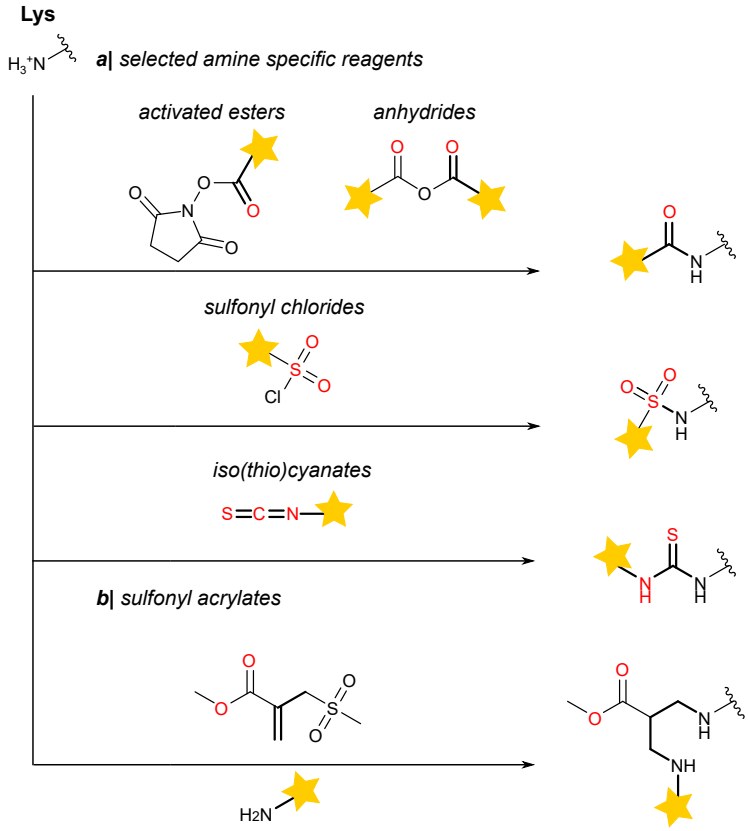


Fig. 1.11: Overview of strategies for Lys modification: **a**| examples of reagents specific for amine group modification[65], **b**| sulfonyl acrylates for site selective Lys modification and subsequent aza-Michael addition with a payload[143].

On the other hand, exquisite site-selectivity can be achieved using engineered substrates, with a prominent example of the SpyTag/SpyCatcher technology[144,145] in which a specific peptide/protein interaction leads to a spontaneous formation of an amide bond between given Asp and Lys residues without the need of carboxylic acid activation. However, proximity based reactivity approaches relying on the molecular recognition of Lys reactive partner (small molecule[146] or peptidic[147]) by the substrate protein are inherently not generally applicable. At times though, the high abundance of Lys residues and non-selective targeting can be an advantage, particularly for high-density cross-linking assays to investigate novel protein-protein interactions[148].

Conclusions

This short overview has highlighted the main strategies for the modification of native residues in proteins, relying on a wide range of chemistries from the repertoire available to an organic chemist. Please note that chemoenzymatic approaches were largely omitted from this discussion (see review by DeForest *et al.* for a good summary of current methods[102]), nevertheless, their impact in specific cases e.g. in vivo settings, should not be neglected. Each selected method comes with its own advantages and disadvantages when it comes to the specific requirements of protein chemistry, which render them useful under different circumstances and for different protein substrates.

It has become more and more apparent that methods based on 'traditional' nucleophilic/electrophilic $2e^-$ chemistry are already stretched to the their limits, for example, the problem of Cys modification is essentially solved. Targeting residues such as Lys, His, Tyr very often inevitably fail to satisfy some criteria, notably the

key chemo- and/or site-selectivity (or even regio-) and create a heterogeneous mixture of products that are often not practical to purify. Strategies based on spatial organisation of reagents to enhance reactivity, while creating powerful solutions to the selectivity issues, are still far from operational simplicity.

Therefore, investigating new types of chemistry, especially targeting less abundant residues in an operationally accessible manner, has been a recent trend in the field. New strategies based on different types of reactivity (metal catalysis, radical chemistry etc.) have brought the possibility of targeting new residues and the extension of structural features that can be manipulated/added. It has to mentioned that many of these key developments are very recent, and have been reported during the course of this research project (2016 - 2020).

In the next part of the this chapter, the applicability of radical chemistry for protein modification as well as the current developments will be discussed in greater detail, with the focus laid on fluoroalkylations.

1.2 Radical chemistry for the editing of biomolecules

Radicals are species with an odd number of electrons, where the non paired electron occupies a single orbital therefore called the SOMO (singly occupied molecular orbital). As such, radicals are reactive intermediates that are essential in modern organic chemistry, having led to profound development in many fields like drug discovery, manufacturing of (agro)chemicals and materials science[149]. Although traditionally regarded as non-selective and uncontrollable, numerous synthetic examples proved that notion wrong[150]. For example, the utility of organotin hydrides in various

transformations such as reductions, deoxygenations and intramolecular cyclisations is well established; so are the useful small molecule synthetic methodologies like Barton decarboxylation, Minisci heterocycle C-H alkylation, Meerwein arylations and many others[151].

The potential of radicals for driving useful chemical transformations is also captured by Nature, as they can be found as common intermediates in biosynthetic pathways such as oxidations with P450 cytochromes[152] or glycyl radical enzyme transformations[153] to name a few. At the same time, some reactive radical oxygen species (hydroxyl), as well as photoinitiated free radicals, are detrimental to the stability of biomolecules in vivo, which has led to the evolution of complex defense systems. However, their role in regulating the stress response, via introduction of PTMs such as Met sulfoxide or side chains carbonyls, is being increasingly recognised, proving the feasibility of this chemistry for biomolecule editing[154].

General features of radical driven transformations seem to be perfectly matched with requirements posed by protein chemistry[151]:

- operationally simple, often under ambient conditions,
- insensitivity towards the solvent, tolerance of aqueous conditions,
- excellent compatibility with many functional groups,
- fast kinetics,
- little influence of sterics.

Considering these properties it is only natural that this type of chemistry could be useful for protein modification, however it is surprisingly rarely exploited for bioconjugation.

1.2.1 Radicals in protein chemistry

The ambiphilic character of radicals allows these species to be simultaneously reactive in an electrophilic and nucleophilic manner, however usually with a different affinity. For example, more electrophilic radicals (e.g. alkoxyl) dominantly react with electron rich sites, while more nucleophilic ones (e.g. C-centred with alkyl substituents) will prefer electron poor sites[155]. Therefore one of the key challenges in protein chemistry - achieving chemoselectivity - strikes as an immediate potential hurdle. Nevertheless, radical chemistry has found many applications and can be encountered in different aspects related with protein editing, which will be briefly discussed below - Fig. 1.12.

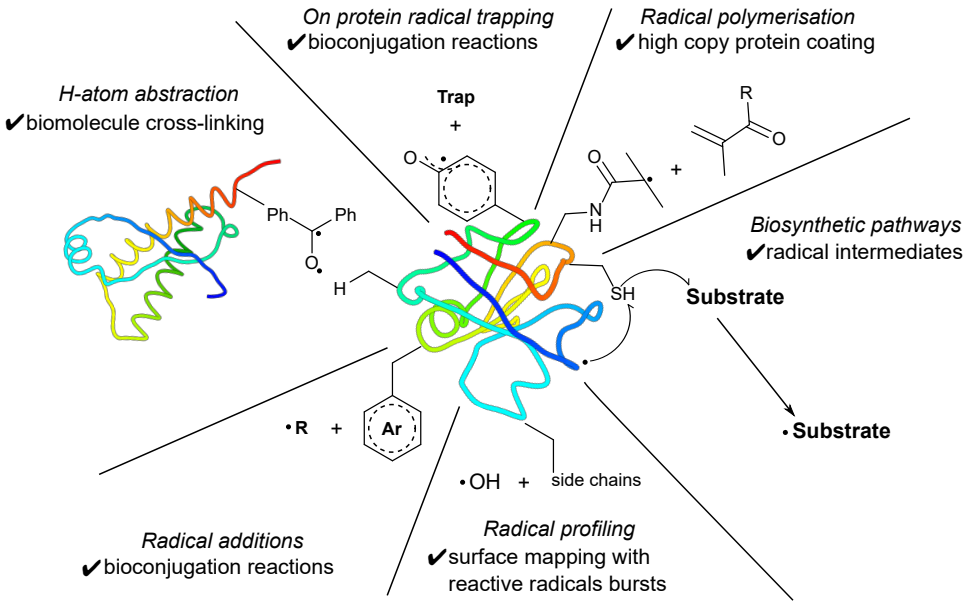


Fig. 1.12: Overview of radical chemistry based applications in protein modifications.

Radicals with engineered proteins - One way to solve the chemoselectivity problem in the specific case, would be to introduce a suitable reactive tag via bioengineering. Indeed, the most prominent example where radical intermediates were used to modify biomolecules was demonstrated for proteins equipped with the Dha tag as a potent SOMOphile[52,53,156]. This strategy provided full chemo- and site-selectivity with the only disadvantage being the lack of the stereocontrol, and hence most probably formation of the diastereomeric mixture of products. Nevertheless, this technology allowed the introduction of a variety of functional side chains - Fig. 1.13 - showing similar reactivity of C-centred radicals with little dependence on the substituents.

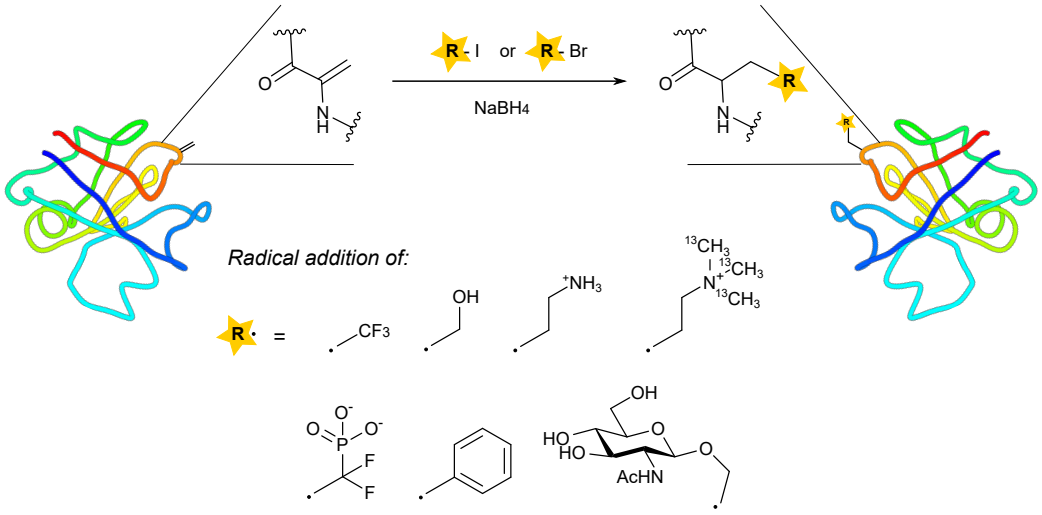


Fig. 1.13: Radical chemistry allowed formal chemical post-translational mutagenesis to a range of residues (only selected examples shown) by reaction of C-centred alkyl radicals with Dha containing proteins[53].

Importantly, an understanding of the reaction mechanism greatly contributed to the improved design of the conditions with use of sodium borohydride instead of Zn metal, leading to reduction in the observed side products of undesired quenching of the long lived capto-dative stabilised C(α) radical (oxygen, multiple addition). Nev-

ertheless, the reaction still requires oxygen free conditions that hinders its operational applicability. Other examples of radical modification of unnatural tags also exploit acceptor reactivity of unsaturated bonds towards S-centred radicals in thiol-ene[157] and thiol–yne[158] reactions. Alkene and alkyne handles can be incorporated using genetic methods and subsequently reacted with the desired thiol in the presence of a radical initiator and UV light. The situation in the absence of a designed radical acceptor is more complicated.

Reactive radical profiling - A substantial focus has been devoted towards modification with extremely reactive species such as hydroxyl radicals (•OH). This interest was spurred by the possibility of fast reaction with many side chains (e.g. Cys, Met, Trp, Phe, Tyr) that has found profound applications in structural studies using techniques like Radical Probe Mass Spectrometry[57]. Such methods can be used to capture information about protein structures and their interactions on a relevant timescale by virtue of oxidation of accessible surface and mass spectrometry readout of digested peptides. An essential feature of these chemistries is the fast generation of radical flux and limited exposure time, to ensure protein integrity. For this purpose, other reactive radicals have also been used, one specific example being laser-induced trifluoromethylation[159]. The instances where oxidation can be applied in a more site-specific manner are limited to, for example, amino acid residues at metal-binding sites where the metal directly catalyses the reactive radical generation[160].

Radical polymerisation - The characteristic feature of radical chemistry, namely the propagation of radicals in a chain reaction, has also been successfully utilised to conduct on-protein polymerisation[161]. This is of particular interest for creating

therapeutic and diagnostic agents for a number of reasons. Coated proteins typically exhibit longer circulation lifetime, lower immunogenicity, allow formulation of higher concentration doses and they can also be engineered to be responsive to a chosen stimuli[162]. The polymerisation reaction requires the presence of an appropriate initiator (e.g. alkyl halide) that is usually attached in a high copy-number (e.g. via Lys residues) and reacted with an olefin monomer. In this manner, protein-initiated grafted-from atom transfer radical polymerisation was used to create a protein-polymer conjugate with superior solubility[163].

Radical mediated cross-linking - A big area of research in protein radical chemistry is the investigation of on-protein radicals in the context of their catalytic properties, inflicted damage, propagation and, less often, synthetic bioconjugation. In the case of amino acids, classical hydrogen transfer reactivity is mainly observed on C(α) (or to a lesser extent next to a stabilising functional group) as a result of the relative stability of these radicals, but has limited synthetic utility[164]. Usually, a C-centred on-protein radicals will dimerise (possibly reaching a sterically accessible position through a series of abstractions, especially to form a thiyl radical) or when in the presence of dissolved oxygen, they are more likely to form a peroxyl species, opening many reactive pathways leading to the protein cleavage, oxidation and formation of carbonyl groups[165]. However, this type of chemistry has found applications in biomolecule cross-linking and subsequent interaction analysis[166]. One chemical moiety heavily utilised is the benzophenone, which upon photoexcitation with near UV light, affords a triplet ketyl biradical with a great affinity towards hydrogen abstraction from C-H bonds, followed by a recombination event, particularly facile with

Gly residues. Therefore, tethering this aryl ketone to a protein provides a platform to form a covalent bond with partners in the vicinity[167].

On-protein radicals - Dimerisation of C-centred radicals is particularly facile for the more long lived tyrosyl radical, which was taken advantage of to perform cross-linking assays[168]. This property offered additional pathways, where the radical could also be trapped with an external radical acceptor to perform bioconjugation reactions (see Fig. 1.5b) in a site-specific manner driven by proximity of the photocatalyst to the protein[169]. The tyrosyl radical could be generated using photoexcited $[Ru(bpy)_3]^{2+}$, or alternatively enzymatically, with horseradish peroxidase and hydrogen peroxide[170].

Using a similar principle - the trapping of an on-protein radical - was utilised in the previously described photoredox C-terminal selective modification[101] of proteins (see Fig. 1.7) and its Trp selective variant where the relatively stable Trp-skatolyl radical could be trapped with Michael acceptors leading to effective C–H activation of the Trp β-position[121]. Both of the methods were demonstrated only for peptidic substrates, presumably due to the requirement of specific conditions; the first achieved the selectivity through low pH of the reaction mixture, the latter inherently gave many side products (e.g. with Lys).

In contrast, Nature has been freely using various radical intermediates to achieve a multitude of transformations. Metal mediated biosynthesis of tyrosyl, tryptophanyl, glycyl and thiyl radicals is used to perform further redox (electron transfer), C-C bond cleavage or O_2 evolution chemistry etc., essential in many metabolic pathways[171].

Radical additions - More discriminatory synthetic transformations can be achieved by utilising the commonly faster addition chemistry (no bond breaking event) and less

reactive radicals (where transition states are late with a substantial radical character on the acceptor site)[172].

Following this idea, several methodologies have been investigated in recent years. Minisci heterocycle modification with alkyl radicals established for small molecules has been applied to peptides[75] and was found to offer good selectivity for protonated His residues - Fig. 1.4e. On the other hand, orthogonal reactivity could be found for the N-centred phenothiazine radicals[92] that primarily targeted Tyr residues - Fig. 1.5g. Finally, certain fluoroalkyl radicals reacted with proteins[123] displayed Trp reactivity - Fig. 1.9d. However, the topic will be critically discussed in the next Chapters of this thesis.

The reactivity of less reactive radicals has not been explored in depth for protein chemistry, with most reports being recent, probably partially due to the lack of analytical tools for unequivocal determination of selectivity and notions of indiscriminatory reactivity. The class of fluoroalkyl radicals investigated for protein chemistry in this thesis will be further introduced.

1.2.2 General features of fluoroalkyl radicals

Properties

Historically, fluorinated radicals were quickly recognised to have different properties from hydrocarbon based radicals, which prompted detailed studies of their structure and chemistry. When treated as substituents fluorine atoms seem to exert mostly electronic effects (as might be expected from the most electronegative element), with steric effects being less pronounced due to the small size. This has a profound effect

on the geometry of fluoroalkyl radicals, as they are primarily pyramidal - Fig. 1.14a, leading to σ-character of the highly substituted species such as CF_3 or CF_2H[173]. In turn, this affects the reactivity of these radicals, as with increasing s-character of the radical species SOMO (lower energy), interactions with the substrates HOMO become more favourable, and hence electrophilicity becomes increasingly pronounced (especially when compared with their hydrocarbon counterparts) - Fig. 1.14b.

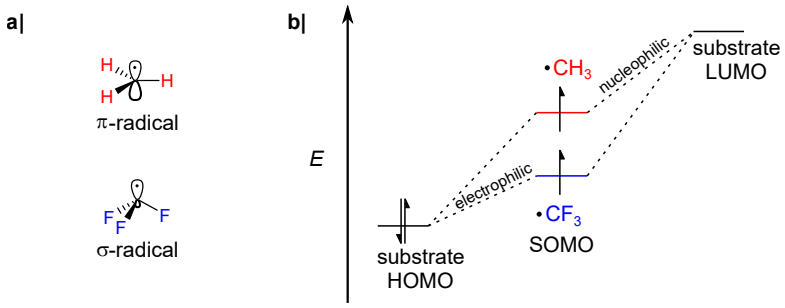


Fig. 1.14: **a**| Comparison of of methyl- and trifluoromethyl- radical geometries, **b**| Simplified analysis of molecular orbital energies for radicals with different SOMO orbitals (methyl 9.8 eV, trifluoromethyl 9.3 eV)[174] and the substrate, showing that the energy gap for (HOMO-SOMO) and (LUMO-SOMO) can be used to explain the electrophilic and nucleophilic character of each of the radical respectively[151].

Therefore, highly fluorinated species are expected to react faster with electron-rich species e.g. alkenes, which was indeed confirmed experimentally in many studies[175].

Generation

Traditional methods for the generation of fluoroalkyl radicals are based on thermal and photochemically activated homolysis of reactive precursors like fluorinated iodides, sulfonyl halides, diacyl peroxides or azoalkanes that have low compatibility with the requirements posed by biomolecules[173]. With conditions restricted to temperatures up to 37 °C, aqueous environment as well as the reluctance to use (protein

damaging) UV light and transition metals, new precursor scaffolds and/or modes of catalysis are necessary to investigate the potential of using these radicals in protein chemistry.

A number of solutions operating under milder, aqueous or aqueous compatible conditions were proposed in recent years involving single electron transfer either in an oxidative or reductive manner, often photocatalytically, and sometimes electrochemically[176]. Notable examples, also from small molecule methodology, operating under such conditions are presented below:

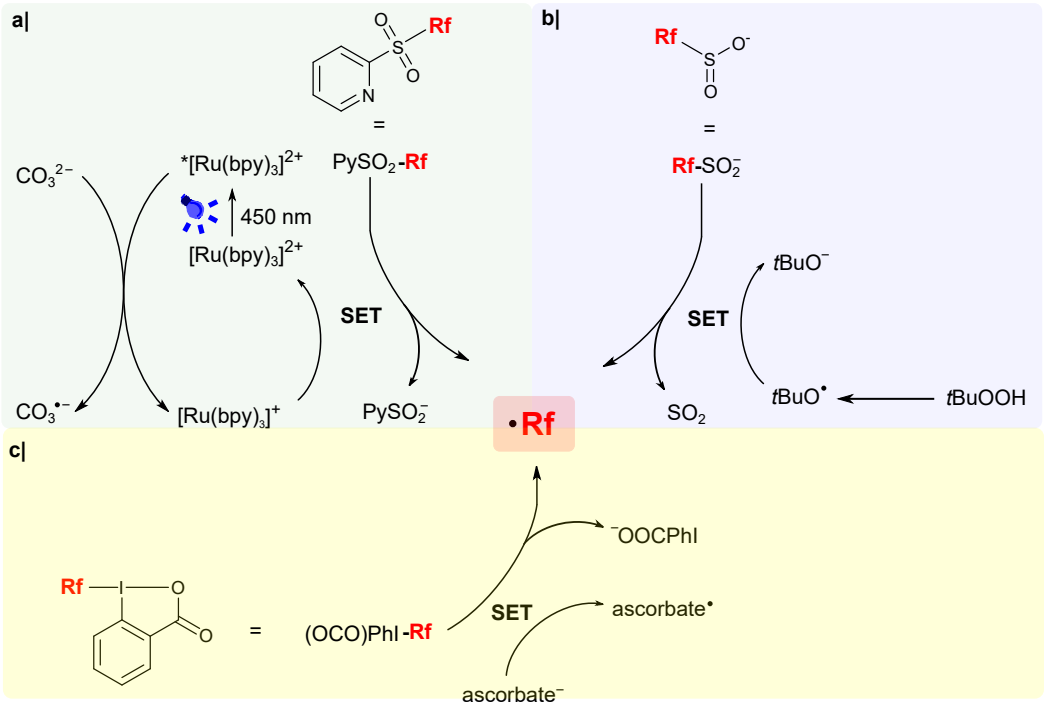


Fig. 1.15: Overview of selected mechanisms for fluoroalkyl radical generation under mild conditions. **a**| $[Ru(bpy)_3]^{2+}$ photocatalytic reduction from pyridyl sulfone[177], **b**| *tert*-butyl hydroperoxide catalysed oxidative pathway from sulfinate salts[178], **c**| ascorbate promoted reductive pathway from hypervalent iodine compounds[123].

(a) Visible light photoredox catalysis allowed room temperature fluoroalkylations of numerous substrates with reductively activated radical precursors[179], e.g. het-

erocycles using sulphonyl chloride and $[Ru(bpy)_3]^{2+}$/blue light[180], fluorinated iodides and zinc phtalocyanine/red light[181] or alkene fluoroalkylation with triphenylphosphonium bromide[182]. The same mode of catalysis was used to activate a more aqueous-compatible radical precursor class of heteroaryl sulfones to achieve isocyanide fluoroalkylations[177] - Fig. 1.15a.

(b) Chemical oxidative catalysis was applied to generate a CF_3 radical from trifluoromethanesulfinate salt in the presence of metals in semi-aqueous conditions[183]. Further development using oxidative activation of that salt was reported to achieve efficient heterocycle modification without metal catalysis[178], while the generality of the methodology and explicit compatibility with buffers was demonstrated also for other fluoroalkyl sulfinates[184] - Fig. 1.15b.

(c) Chemical reductive activation provided fluoroalkyl radicals[185] from hypervalent iodine compounds in a transition metal catalysed manner[186] but was also refined to metal free conditions[187] - Fig. 1.15c. Eventually, the technology was applied under truly biocompatible conditions with sodium ascorbate as a reductant[123] (see Fig. 1.9d).

Reactivity

The principal reactivity of fluoroalkyl radicals is the addition to unsaturated systems (usually several times faster than H-abstraction), with another useful pathway of substitution reaction based on the $S_{RN}1$ mechanism (Unimolecular Radical Nucleophilic Substitution) for perfluorinated halides[173].

This has been thoroughly exploited in small molecule synthesis methodologies[150,179].

Interestingly, from the perspective of protein chemistry reactions on compounds resembling side chains of Trp, His, Phe, Tyr and Cys have been reported.

Indoles - Being a medicinally relevant heterocycle, a number of methodologies have been developed for indole fluoroalkylations. The regioselectivity of the radical addition depends on the substitution pattern, especially in the nitrogen containing ring. With the C3 position blocked, as for Trp, the dominant modification occurs on C2[188]. The agreed upon mechanism for this transformation is shown in Fig. 1.16 and it starts with addition of the free radical to C2, generating a stabilised radical on C3. In a typical course of the reaction, the resultant species is oxidised by a single electron transfer to the carbocation that then aromatises upon proton abstraction - Fig. 1.16a.

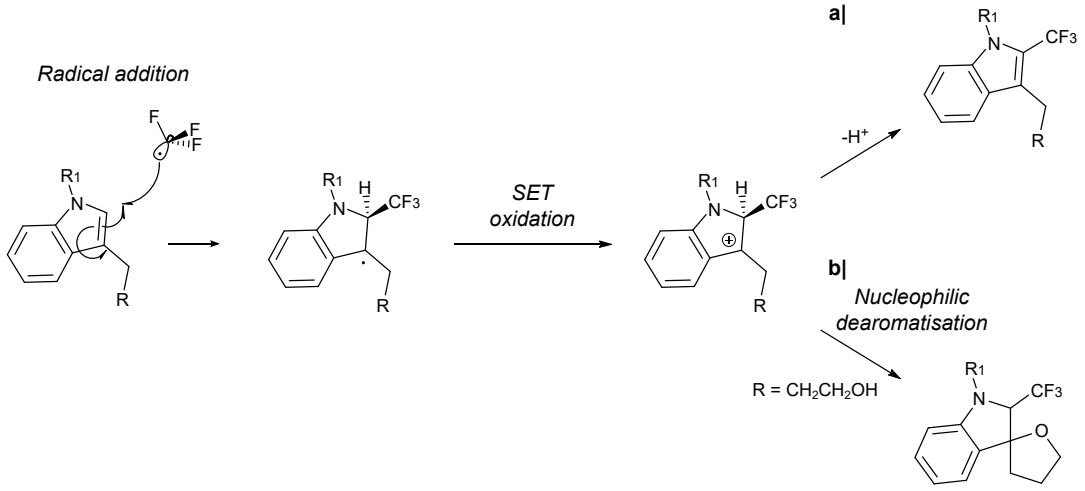


Fig. 1.16: Mechanism of free radical addition to C3-substituted indoles - an example of oxidative trifluoromethylation[189].

However, in the presence of a suitable internal nucleophile, the intermediate carbocation can be trapped to form a spiropyrrolidinoindoline[189] - Fig. 1.16b.

Imidazoles - Similar attention was devoted to the modification of imidazoles with

many available strategies[190]. Fluoroalkylations of unsubstituted imidazoles tend to give a mixture of products with some formed in an excess[191]. Similarly, initial results for trifluoromethylation of His residues suggested that the radical addition leads to the formation of a mixture of isomers modified at both available positions[192]. More recent results on a wider range of radicals suggest higher selectivity for C2 modification[75,78].

Phenols/phenyls - Simple aromatic systems with various substituents can also be efficiently modified by the fluoroalkyl radicals[176,179]. It does not come as a surprise that groups with the electron donating character (e.g. hydroxyl group) provide higher reactivity of the ring. In the case of a single alkyl substituent (Phe analogue), almost an equal mixture of all possible isomers were formed with trifluoromethylation[193], while *para*-substituted phenols were modified primarily at positions *ortho* to the hydroxyl group, with a significant amount of multiple addition products also formed[183,194].

Thiols/disulfides - Trifluoromethyl radicals have been used for disulfide splitting in various substrates, including amino acids[195], leading to the formation of the pharmaceutically relevant SCF_3 group. A similar transformation was proposed for free thiols and thiolates using a range of conditions and radical precursors[196]. Not surprisingly, the methodology could also be extended to perfluoroalkylations[197].

In summary, the reactive side chains of native residues and probable positions for fluoroalkyl radicals attack are shown in Fig. 1.17.

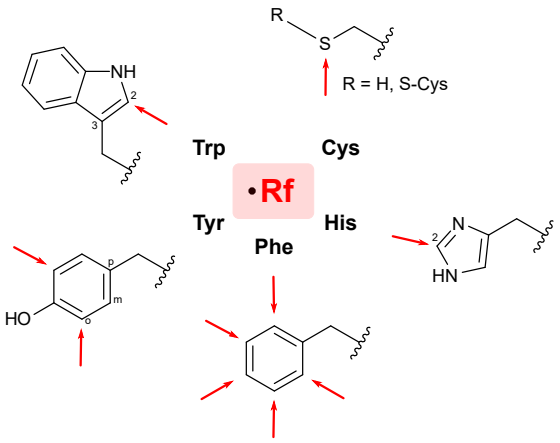


Fig. 1.17: Reactive AA side chains and probable positions for a fluoroalkyl radical attack[179].

Conclusions

Radical chemistry has proven to be compatible with proteins, creating useful constructs, and delivering unique tools for relevant studies in chemical biology. However, the review of the available literature clearly shows it is underutilised for synthetic transformations on native proteins. At the beginning of this research project, there were no comprehensive studies on the radical fluoroalkylations of complex peptides, yet alone proteins. As such, the outcome of such transformations with respect to both reactivity and selectivity was unknown at the start of the project. The idea to perform trifluoromethylation on proteins was proposed in one report[198] dealing with Cys electrophilic trifluoromethylation in peptidic substrates. In this report, a competitive reactivity of Trp residues was observed, but in-depth studies were not performed. As the authors themselves admitted, the study should be treated preliminary, awaiting optimisation, scaling up and demonstration of applications.

The last part of the introduction will look at current methods for the generation

of fluorinated proteins as well as their general utility and applications in the field of chemical biology.

1.3 Fluorination of proteins

An additional motivation for the development of a chemical strategy for protein modification based on fluoroalkyl radicals can come from the fact of fluorine introduction. Interestingly, considering its relatively high abundance as an element, it is rarely encountered in biological settings[199]. However, fluorine's unique properties can have a profound impact on the physicochemical properties of the compounds, example being regulation of neighbouring groups pK_a and impact on hydrophobicity - effects that have been readily exploited in medicinal chemistry[200] as well as other fields[201]. These benefits have also been utilised in the field of chemical biology for the development of small molecule reagents such as uniquely reactive probes for activity-based protein profiling, affinity labelling, protein foot-printing and cross-linking[202].

Similarly, when introduced into biomolecules, fluorine can provide a range of applications both as an analytical probe[203] and a structural component[204]. This methodology has been gaining momentum with the 're-discovery' of protein observed ^{19}F NMR (PrOF) e.g. for studying enzymatic mechanisms[26,27] and has been finding new applications in engineering and influencing protein activity/stability[205,206]. The next section will deal with methods for the generation of fluorinated proteins.

1.3.1 Generation strategies

There are two main approaches for constructing proteins containing fluorinated residues: biosynthetic, where fluorinated residues are incorporated at the time of expression, and chemical, in which the protein is decorated with fluorine groups after the expression takes place.

Biosynthetic methods

There are several ways to bio-incorporate fluorinated amino acids into proteins, the choice of which often depends on the aim of the study[207]. If a precisely positioned label is needed, then genetic code expansion provides the best results. A stop codon can be inserted into a desired position and its natural function suppressed by a designed tRNA carrying the desired fluorinated amino acid analogue, resulting in its site specific incorporation during translation[208]. Though powerful, its use requires substantial effort towards developing efficient orthogonal tRNA/aminoacyl-tRNA synthetase pairs. Notably, this technology was used to install a single Phe-CF_3 residue which was subsequently used as a protein structure and reactivity probe in vivo[209]. Examples of other incorporated structures are shown in Fig. 1.18a. On the other hand, if a higher amounts of fluorinated protein residues are needed, this technique becomes impractical and the global substitution of a residue is preferred. Such methods can successfully yield highly fluorinated proteins and have been commonly practised, also due to their relative operational simplicity[204].

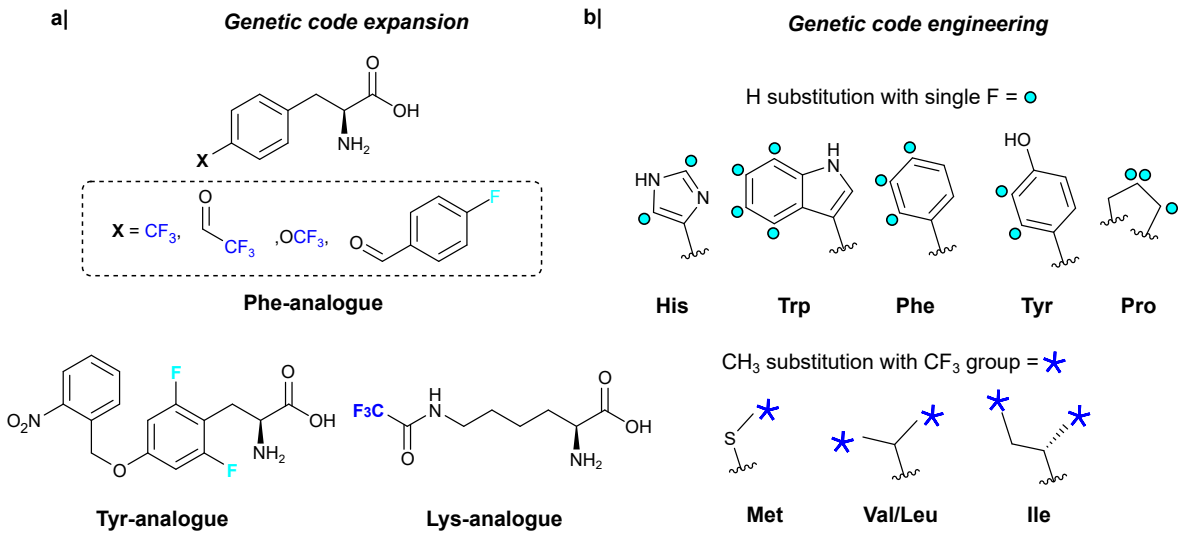


Fig. 1.18: Selected examples of biosynthetically accessible residues for protein fluorination[207]. **a**| Selected amino acid structures that can be incorporated using stop codon reassignment technology, **b**| Amino acids that can be substituted with fluorinated analogues for residue specific incorporation.

Since hydrogen to fluorine substitution is generally non-perturbing, endogenous cell machinery can be used for ribosomal synthesis of a modified protein in an auxotrophic host cell supplemented with the fluorinated amino acid analogue. The main constraint of this method is the repertoire of fluorinated AA analogues that is limited by the similarity to their native counterpart. Nevertheless, fluorinated proteins accessed using this methodology have been used to study, among other examples, enzyme dynamics[210]. Typically targeted residues are shown in Fig. 1.18b.

Chemical methods

Although biosynthetic methods can be applied universally, it is only via chemical methods that ^{18}F-labelling can be realised due to the short half-life ($\approx$ 110 min) of this radioisotope. In this case, the fluorine introduction can also be achieved either in a site- or residue-selective manner. For the first, it is possible to chemically position a

label in a site-specific manner using engineered proteins via 'Tag and Modify' strategy. In practice, this is most often achieved via site-directed mutagenesis to introduce a Cys residue in the position of interest and careful alkylation with a fluorine containing prosthetic group (alkyl halide)[211]. An additional reason for the success of this type of strategy is the commercial availability of the reagents - Fig. 1.19a. Robustness was demonstrated when this method could be used to characterise the conformational landscape of a cell surface receptor[212].

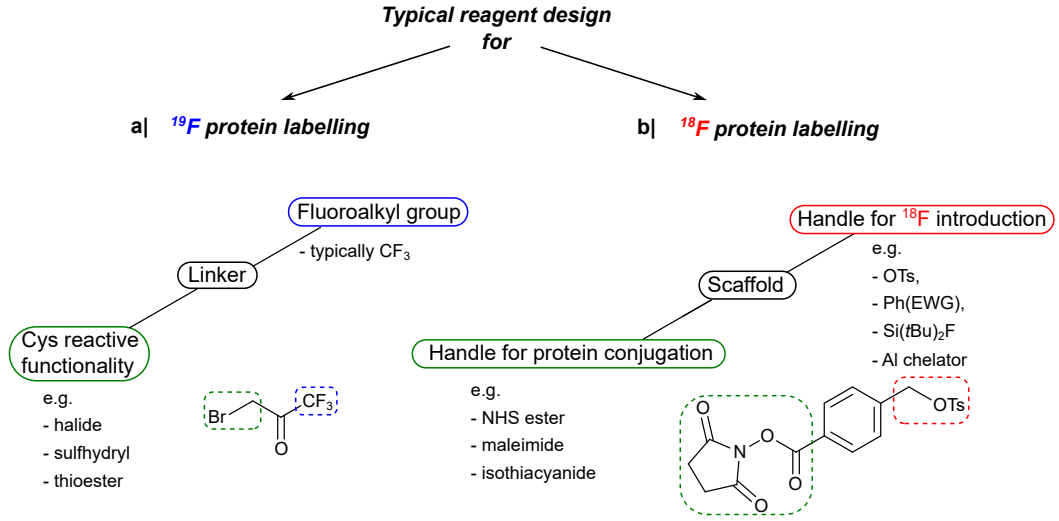


Fig. 1.19: **a**| Analysis of a typical (commercially available) reagent for introduction of ^{19}F tag into proteins[211], **b**| Analysis of a typical reagent for introduction of the ^{18}F tag into proteins[213].

In principle, any other method (e.g. discussed in section 1.1.2) that achieves site-specificity, can be used, provided that the prosthetic group is equipped with fluorine. In practice, such strategies are rarely employed for analytical purposes (NMR) as the necessity of a lengthy linker can significantly perturb the protein structure and render the studies less relevant. On the other hand, this seems to be of lesser importance when it comes to ^{18}F-labelling, as it is often achieved via residue specific

chemistry (Lys, Cys). This strategy is particularly popular for the antibody radiolabelling, where high reactivity and fast kinetics are the prevailing factors[213]. Typical construction of the reagents is presented in Fig. 1.19b. Sometimes ^{18}F-labelling can be achieved with readily available derivatives of radiolabelled glucose, for example by reaction with Cys/Dha tagged protein[214]. On the other hand, apart from introducing a prosthetic with the [^{18}F]C-F bond, radiofluorination can be achieved via direct boron-fluorine and silicon-fluorine bond formation or via chelation with aluminum complexes, requiring corresponding functionalities to be attached to the protein[215].

Highly fluorinated proteins are most efficiently obtained through chemical synthesis, which although laborious and limited in protein size, allows a complete freedom in the design[216]. It is worth noting the technical progress that has been achieved in this area, for example by the application of automated flow chemistry[217], holds promise for a more widespread utilisation of protein synthesis in coming years.

1.3.2 Applications

Broadly speaking, the applications of fluorinated proteins can be grouped into two categories, analytical and structural. General overview of the most typical use is presented in Fig. 1.20 and briefly discussed below.

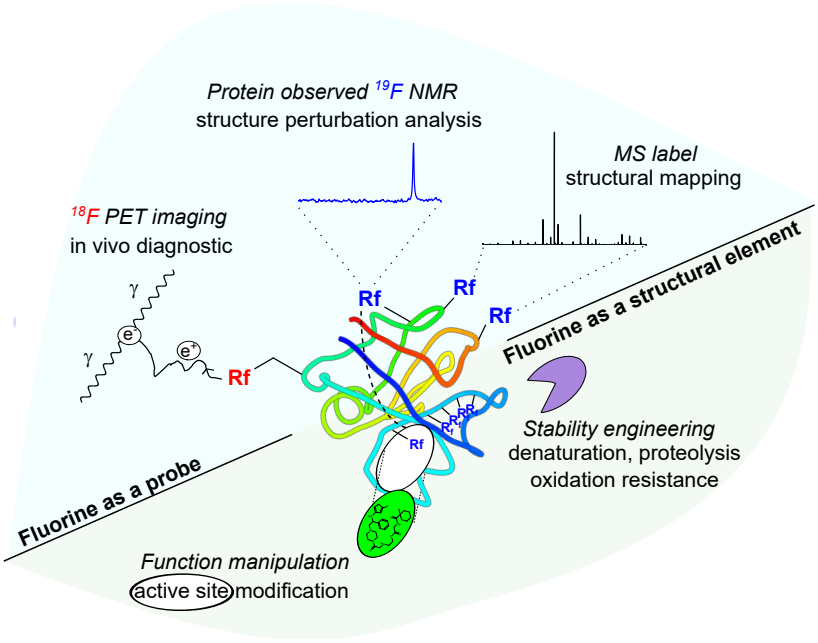


Fig. 1.20: Overview of fluorinated protein applications highlighting the role of fluorinated residues as analytical probes[203,213] or structural elements[218,219].

Fluorine as a probe

One of the most long standing[220] and recognised applications for proteins with fluorinated labels is their use in protein observed ^{19}F NMR. The particular characteristics of the fluorine nuclei provide great advantages in these types of experiments[203]:

- background free signals (fluorine is absent from biology),
- high sensitivity of the chemical shift to the local environment (structure perturbation analysis),
- good experimental sensitivity, standard instrumentation, 100% natural abundance of the NMR active isotope (similar nuclear properties to proton: gyromagnetic ratio, spin of 1/2),

- large chemical shift dispersion (prevents signals overlap),
- large dipolar coupling (activity in nOe (nuclear Overhauser effect) based experiments),
- relatively small van der Waals radius (minimal impact labels)[221].

Therefore, it does not come as a surprise that fluorinated proteins have been used in a wide range of NMR experiments. Although, ^{19}F NMR analysis of small molecule ligands is already an established technique for fragments/hits screening, it is obviously constrained to fluorinated molecules[222]. However, the same technique can be applied for *any* potential ligand if analysed from the perspective of the protein. Thanks to perturbations in the protein folding, reflected in the chemical shift change, the thermodynamics of the interaction can be quantified[223]. As mentioned before, most prominent examples are based on investigating peak shape/position and their relationships with a conformational change of the protein, occurring in a variety of biological processes like (mis)folding, aggregation, protein-partner interactions, etc.[203,210,212]. The resolution of data could be further enhanced by employing additional ^{13}C-labelling and novel NMR experiments, demonstrating an ongoing need for labelling techniques[224].

Although suffering from broadened signals, the experiment is also amenable to in cell studies, with the potential to provide information about the proteins in their natural environments[225].

The other important application of fluorinated proteins is their use in clinical diagnostics upon labelling with the radioactive isotope - ^{18}F, followed by the PET

(positron emission tomography) imaging[226]. This isotope is typically produced in a cyclotron, upon proton bombardment of $[^{18}O]H_2O$, as a fluoride ion and then incorporated into a protein using a prosthetic group as discussed previously. The nuclei are unstable and upon the decay they release a positron which recombining with a nearby electron produces two γ rays that can be detected via the PET camera. Since both photons are released along a parallel axis in an opposite direction, coincidence detection and 3D reconstruction allows creation of an image[227]. This mode of imaging is especially valued in in vivo studies, as it is non-invasive and requires trace amounts of material, avoiding potential toxicity issues. Therefore, radiofluorination of antibodies and their fragments, with affinity for molecular targets implicated with a given disease, has been increasingly popular[213].

One can also note that fluorinated groups can serve as a mass spectrometry label[159], as discussed briefly earlier, although the advantage of fluorine here would lie from reactivity perspective rather than a unique detection mode.

Fluorine as a structural element

The other major application is the use of fluorine to manipulate structural features of proteins. Perfluorocarbons, in addition to being chemically nonreactive, are also particularly hydrophobic, which allow for the rational substitution of amino acids of this character to their fluorinated analogues. Crystallographic studies have found that such a change tends to preserve the side-chain's shape with a minimal increase in volume, which usually conserves the fold and strengthens the hydrophobic effect[218]. Therefore, the overall impact results in stabilisation of the protein structure, how-

ever, the exact effect seems to be poorly predictable at the moment. Particularly good results were achieved when targeting a defined α-helical secondary structure, while replacing other amino acids in the hydrophobic core of proteins was shown to cause slight destabilisation[204]. The potential to increase resistance of enzymes (to denaturation, proteolysis, oxidative damage etc.) by fluorination could be of particular interest in industrial or clinical settings.

On the other hand, fluorinated residues can also influence the catalytic performance of proteins when positioned in the active site. As such, they can be used to identify of functional residues (e.g. upon changes in their properties like nucleophilicity)[219]. In one example, a fluorinated analogue could restore activity of a mutant protease inhibitor, effectively serving as a mimic of a polar residue[205].

Conclusions

Fluorinated proteins proved to be increasingly impactful in various chemical biology experiments. However, their widespread use seems to be hampered by the scarcity/availability of strategies for fluorine introduction. Biosynthetic methods can deliver proteins suitable for most current applications, with a notable exception of radiolabelling, and therefore it is beneficial to develop possible complementary chemical strategies that can be used for ^{18}F introduction.

Interestingly, from the point of view of protein observed ^{19}F NMR studies, observation of fluorinated aromatic residues is especially noteworthy. Their relatively low abundance and widespread distribution in the hydrophobic regions of proteins is often beneficial to observe and analyse ligand interactions[228]. However, there are no

chemical methods for specific *direct* fluorination of these residues with minimal size tags.

1.4 Aims of this thesis

As introduced above, the chemical modification of proteins provides enabling tools for studying biology. Exquisite solutions have been proposed for the labelling of engineered proteins using 'Tag and Modify' strategy. Methods for 'classical' bioconjugation chemistry of native proteins have been greatly optimised in recent years. An ever growing number of strategies makes it possible to target novel residues, however, the need for robust, generally applicable and operationally simple methods still remains. Focusing our attention on radical chemistry, which may fulfill these criteria, we sought to investigate the possibility to advance the field through novel applications to native proteins. Specifically, we chose fluoroalkyl radicals, which apart from yielding a new bioconjugation strategy could provide a new way to introduce fluorine into proteins. A review of the literature, revealed their favourable reactivity patterns and a promising progression in biocompatible radical generation. Strikingly, there have been few studies about their selectivity in peptide/protein chemistry. Therefore, the specific aims that are to be addressed in this thesis include:

1. assessment of the reactivity and selectivity for chosen fluoroalkyl radicals for protein modification,

2. investigation of developed methodologies for adaptation to ^{18}F radiochemistry,

3. evaluation of the methodology as a potentially generally applicable bioconjuga-

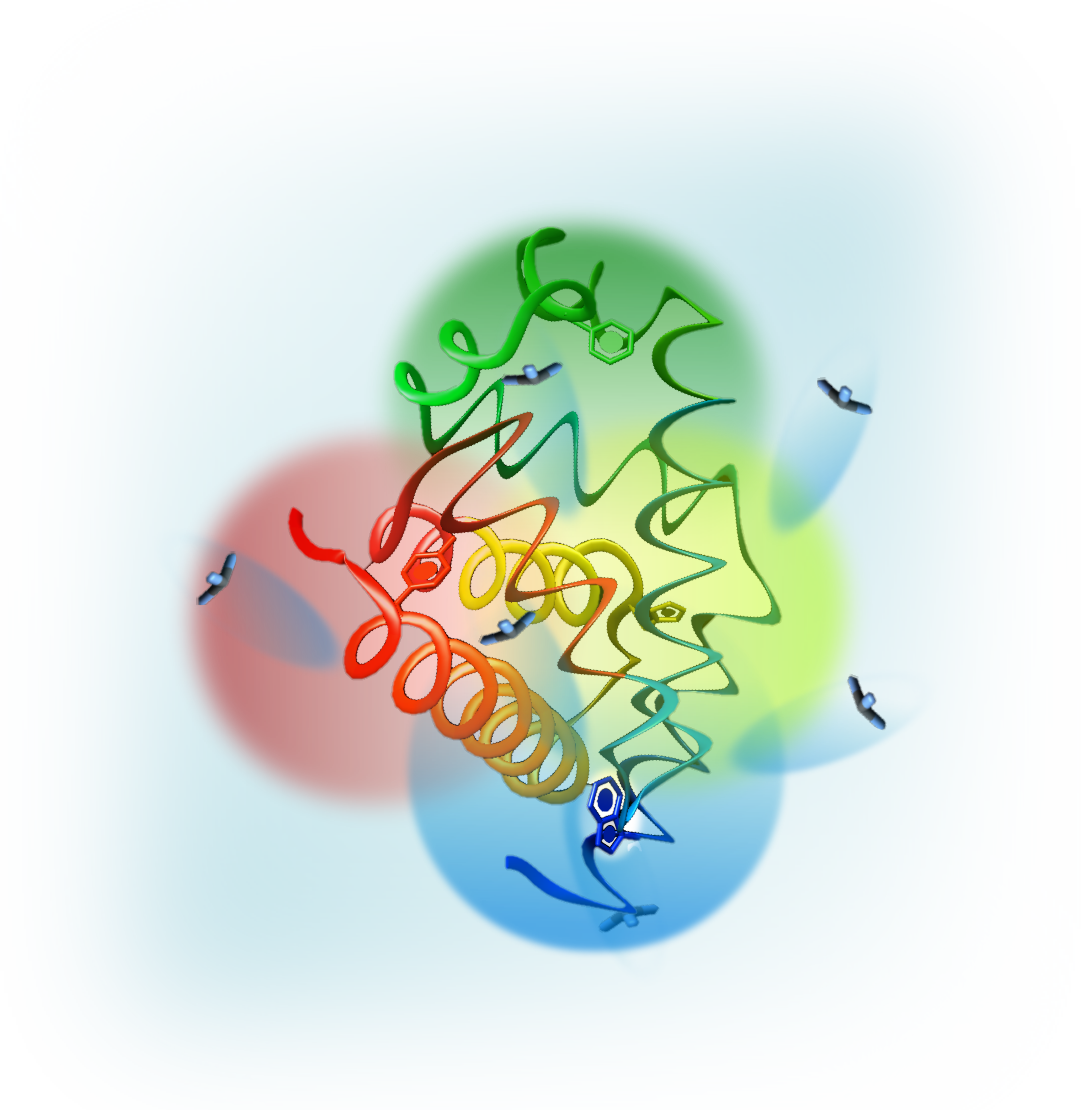
tion strategy for native aromatic residues modification,

4. characterisation and exploration of possible applications of prepared fluorinated constructs, especially in relevant fields like protein observed ^{19}F NMR.

The aims of this thesis fit the current challenges in protein chemistry, where exploration of new methodologies (here radical) is envisioned to push the frontiers and open up unprecedented applications[229].

Chapter 2

Selective trifluoromethylation of native residues in proteins



Abstract

Chapter 2 initiates the research on radical protein fluoroalkylation by the study of trifluoromethylation - a transformation opening applications in protein observed ^{19}F NMR. Trifluoromethyl sulfinate was demonstrated to be a protein compatible radical precursor following oxidative initiation. Detailed studies of this system were conducted to understand the reaction characteristics using substrates of several levels of complexity. These model systems revealed the possibility of chemoselective targeting of one amino acid side chain over others - namely the indole motif of tryptophan. Selectivity conclusions coming from these studies were evaluated using a range of techniques and tested in many case-studies revealing a more complex picture, in which the specific protein (hetero)aromatic content may dictate the fine ratios of the products. Nevertheless, the prepared constructs, containing a uniquely positioned chemically introduced trifluoromethyl groups, were amenable to protein observed ^{19}F NMR studies, allowing access to sensitive assays for the elucidation of ligand binding and structural changes.

2.1 Introduction

To investigate the radical fluoroalkylation of native residues in proteins, trifluoromethylation was chosen as the first project. From the reactivity perspective, many precedents in the literature suggested high reactivity with (hetero)aromatic residues, as discussed in the previous chapter[183]. Additionally, one may speculate that the fairly small trifluoromethyl radical may not be influenced by the steric hindrance, and hence will provide a comprehensive reactivity profile independent of the congested environment of protein residues. Furthermore, a CF_3 tag is preferred for protein NMR applications, providing a superior sensitivity multifluorine system. It also displays favourable chemical shift anisotropy properties, which leads to a more narrow fluorine NMR signal, particularly helpful in the study of larger biomolecular complexes[230]. Finally, there are variety of established generation methods, which should provide ample screening opportunities[231]. Therefore, trifluoromethylation was chosen as both a methodology that can be used to establish and evaluate the workflow, as well as one that can provide attractive applications.

At the beginning of this project, there were few reports on the trifluoromethylation of complex biomolecules, including preliminary studies for peptides[198] or non-selective radical profiling[159] (see Chapter 1). However, after the publication of the results of this project[232], several new contributions appeared (predominantly dealing with peptide modifications)[123,194,233,234], which are discussed at the end of this chapter.

The problem of fluoroalkylation in aqueous media has been considered in Chapter 1.2.2. Overwhelming number of studies on *radical* trifluoromethylation have been

performed using trifluoromethyl sulfinate (as a sodium salt - NaTFMS), initiated with an organic oxidant[176] - and the general mechanism was discussed in detail in Fig. 1.15b. Moreover, trifluoromethylation of small molecules with this reagent was reported to be successful under protein biocompatible conditions, and importantly, did not impede the function of a protein that was present in the buffered aqueous reaction mixture[184]. From a broader perspective, a system based on sulfinates can also be used to perform other fluoroalkylations, as the radical generation event was reported to be universal for various fluoroalkyl groups. Additional desirable features of this system are the commercial availability, excellent solubility in aqueous conditions, and many adaptable redox initiation modes (photocatalytic etc.)[235]. Accordingly, trifluoromethyl sulfinate/oxidant system was chosen as the platform for generating CF_3 radicals.

The results are presented starting from the small molecule reactivity confirmation, along with an assessment of selectivity. Next, the reaction was translated towards peptide modification, and then studied extensively using a model protein. Finally, the methodology was applied to further protein substrates for which proof-of-concept protein observed fluorine NMR experiments were conducted[232].

2.2 Small molecule and peptide model

2.2.1 Reactivity

The investigation started with an assessment of the reactivity using model small molecules - unprotected amino acids, reacted with an excess of reagents in a fully aqueous buffered solution - typical conditions used for protein chemistry. For the

screening, ^{19}F NMR was utilised, taking advantage of the characteristic chemical shift of the trifluoromethyl group attached to the aromatic system (ranging from −52 to −87 ppm)[236]. To identify the side products, first, a 'blank' reaction was investigated, which revealed formation of several new species - Fig. 2.1a.

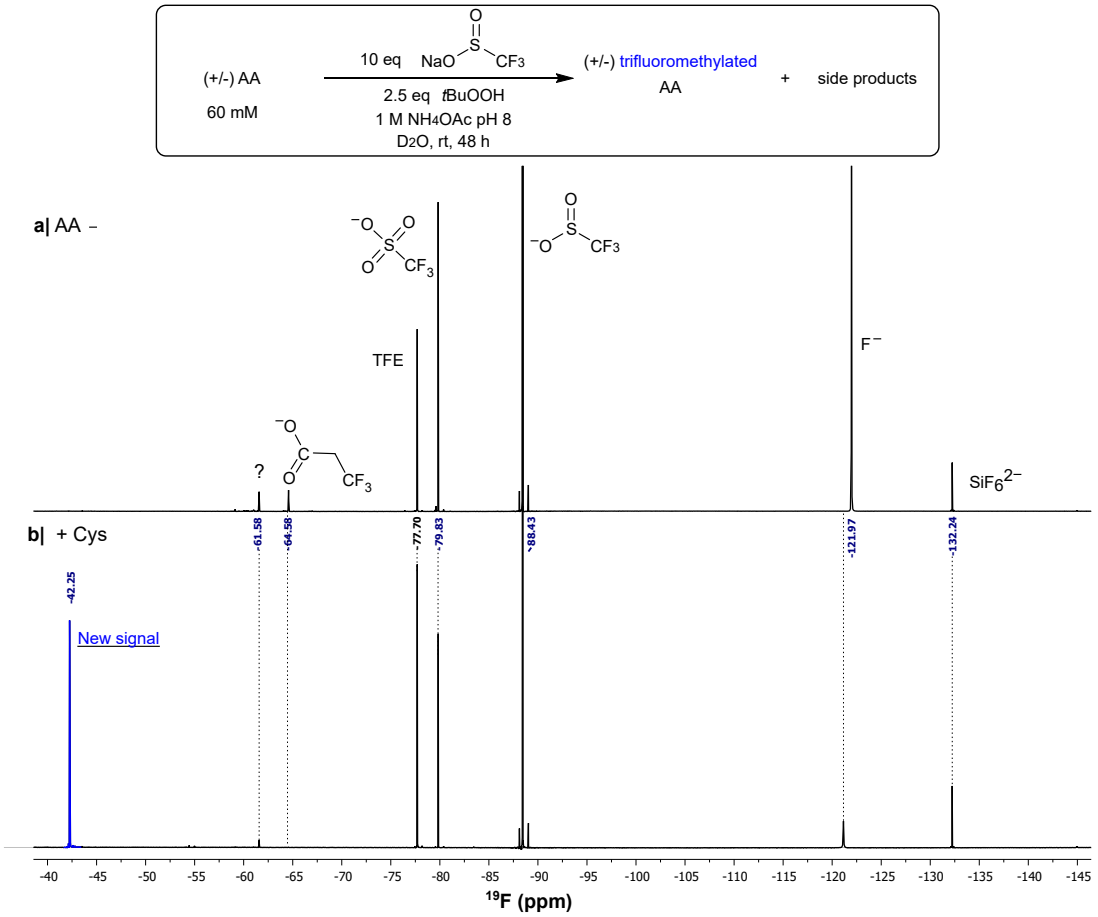


Fig. 2.1: ^{19}F NMR analysis of amino acids trifluoromethylation. **a**| Spectrum of the reaction mixture without the AA substrate with peaks assignment. **b**| Spectrum of reaction mixture with Cys as the substrate. 2,2,2-Trifluoroethanol (TFE, −77.7 ppm) was used as an internal standard.

Starting downfield; two, relatively weak but distinct triplets were identified as the most likely products of a background reaction with acetate from the buffer. One (at −64.6 ppm), could be identified as 3,3,3-trifluoropropanoic acid by comparison of its

^{1}H and ^{19}F chemical shifts and coupling constants with those reported in the literature[237]. These signals were absent when the experiment was repeated in a reaction mixture without the ammonium acetate (see Fig. 2.2 blank−). Continuing upfield, the peak at −79.8 ppm was attributed to the oxidised derivative of the NaTFMS - triflate, while the next one at −88.4 ppm was the reagent itself[236]. Shielded peaks at −122 pm and −132.2 ppm (with visible ^{29}Si satellites) may indicate the presence of inorganic fluorides (in the different pH dependent forms) and other side products from a reaction with the glass[238]. Such an assignment allowed ready identification of the peaks coming from the reaction with the amino acids, as shown in the case of Cys - Fig. 2.1b.

Using this ^{19}F screening method, the reaction was repeated for all the native AAs, all residues yielding prominent products were identified - Fig. 2.2 (corresponding to >1% conversion as integrated versus the internal standard). As expected, only amino acids with (hetero)aromatic side chains (as well as free Cys) yielded a distinct set of singlet signals. Immediately evident was the superior reactivity of the Trp, His and Cys residues, reaching >20% relative conversion. The other amino acids yielded only the common side products present in the blank reaction (−61.6 ppm and −64.6 ppm). In the cases of His, Phe and Tyr, the observed signal pattern resulted from the formation of multiple regioisomers as well as double addition products as confirmed by LC-MS analysis. Only Trp and Cys gave close to a homogeneously modified fluorinated product, prompting a more detailed investigation of these species.

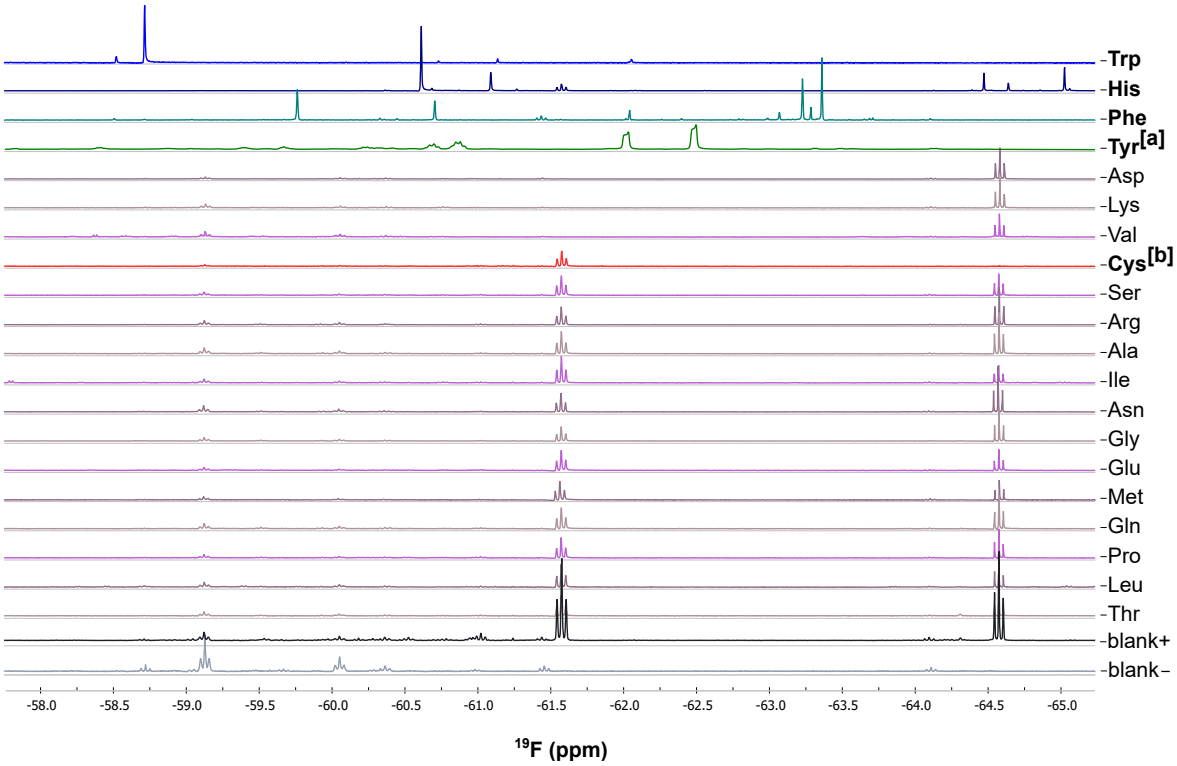
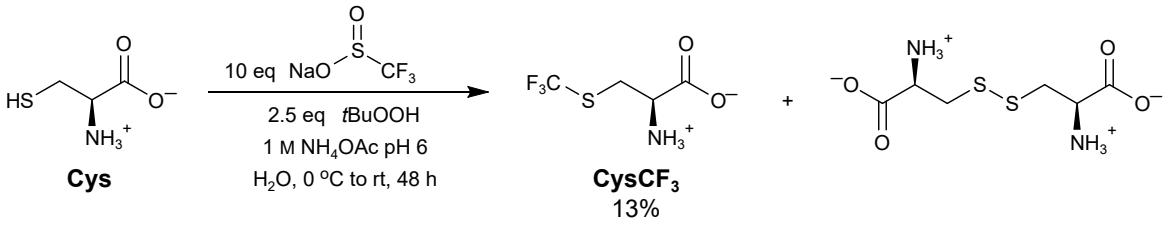


Fig. 2.2: Fragments of ^{19}F NMR spectra recorded after 48 h of reaction for the unprotected AAs and controls (blank−: no AA, non-buffered conditions, blank+: no AA, 1 M ammonium acetate pH 8). [a] Tyr was not fully soluble under the reaction conditions (with a solubility of 1.3 mg/mL (∼13%) by the quantitative ^{1}H NMR analysis). [b]Main peak for the reaction with Cys was shown in Fig. 2.1b.

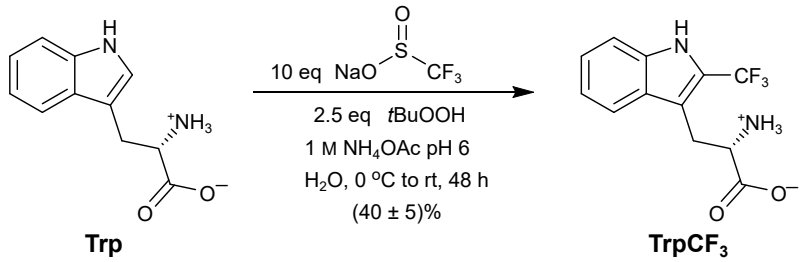
Reaction with Cys turned out to yield the *S*-trifluoromethylated compound - **CysCF$_3$**, but also predominantly formed an abundant white precipitate, subsequently identified as an insoluble oxidation product - cystine - Sch. 2.1.



Sch. 2.1: Radical trifluoromethylation of cysteine (30 μmol, 50 mM).

On the other hand, trifluoromethylation of Trp led to the isolation of C2 derivative -

TrpCF$_3$, in a much better yield - Sch. 2.2. The fluorine chemical shift of this compound (−58.7 ppm) matched the most prominent from the crude reaction mixture.



Sch. 2.2: Radical trifluoromethylation of tryptophan (90 μmol, 50 mM, $n = 3$).

This preliminary data suggested that the proposed trifluoromethylation protocol may be suitable for a chemoselective Trp reaction, therefore further investigation of the various reaction parameters was performed for this AA. The reaction was repeated under a variety of conditions and the conversions after 48 h were determined via quantitative ^{19}F NMR analysis - Tab. 2.1.

Tab. 2.1: Screening of different conditions for the Trp trifluoromethylation. In all cases 2.5 eq of *t*BuOOH was used on a 30 μmol AA scale at concentration of 50 mM.

Entry	**Buffer/pH**	**NaTFMS [eq]**	**Additional conditions**	**Conversion [%]**
1	D_2O final pD ∼2	10	-	35
2	100 mM NH_4OAc pH 8 final pH ∼5	10	-	49
3	1 M NH_4OAc pH 4	7.5	-	27
4	1 M NH_4OAc pH 6	10	-	56
5	1 M NH_4OAc pH 6	10	500 rpm stirring	54
6	1 M NH_4OAc pH 6	10	5 eq $ZnCl_2$	50
7	1 M NH_4OAc pH 8	7.5	-	36
8	1 M NH_4OAc pH 12	10	-	0

The main conclusions from these experiments are summarised below:

- If the reaction was performed under the non- or weakly-buffered conditions (entries 1 and 2), a rapid acidification was observed, which is in agreement with the proposed mechanism[178] - as discussed in Fig. 1.15b.

- The highest conversions, reaching around 50% after 48 h, were achieved in neutral pH, buffered reaction mixtures (entries 2, 4, 5 and 6).

- Stirring seemed to have little or no effect on the observed conversion (entry 5). As a standard procedure, the reaction mixtures were not stirred as it was previously found to give detrimental results[178].

- Established protocols highlight role of zinc salt in achieving an efficient transformation[184]. However, here the presence of $ZnCl_2$ (entry 6) did not affect the observed conversion.

In the next step, ^{19}F NMR was employed to monitor the conversion to the **TrpCF$_3$** product under different conditions - Fig. 2.3.

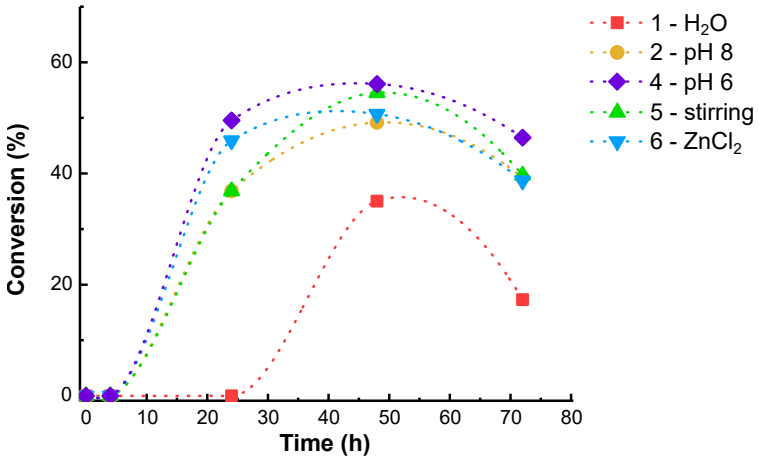


Fig. 2.3: ^{19}F NMR time-course of Trp trifluoromethylation, with quantification of the conversion to the main product at −58.6 ppm, under different conditions (see entries in Tab. 2.1 for details), dotted lines show apparent fit.

Most of the reactions under buffered conditions had a similar profile, confirming the small influence of stirring or the presence of Zn^{2+} (green and blue). Interestingly, all of the reactions were characterised by an induction period (no product after the first 4 h and an even longer time for the reaction under non-buffered conditions) with the maximum conversion limited by further reaction of the trifluoromethylated product. This was indicated by a decrease in the amount of product post 48 h, most likely due to double/multiple CF_3 additions as supported by an increase in the number of signals in the (−50 to −65) ppm region.

In order to further investigate side products formation an extended NMR analysis was carried out - Fig. 2.4.

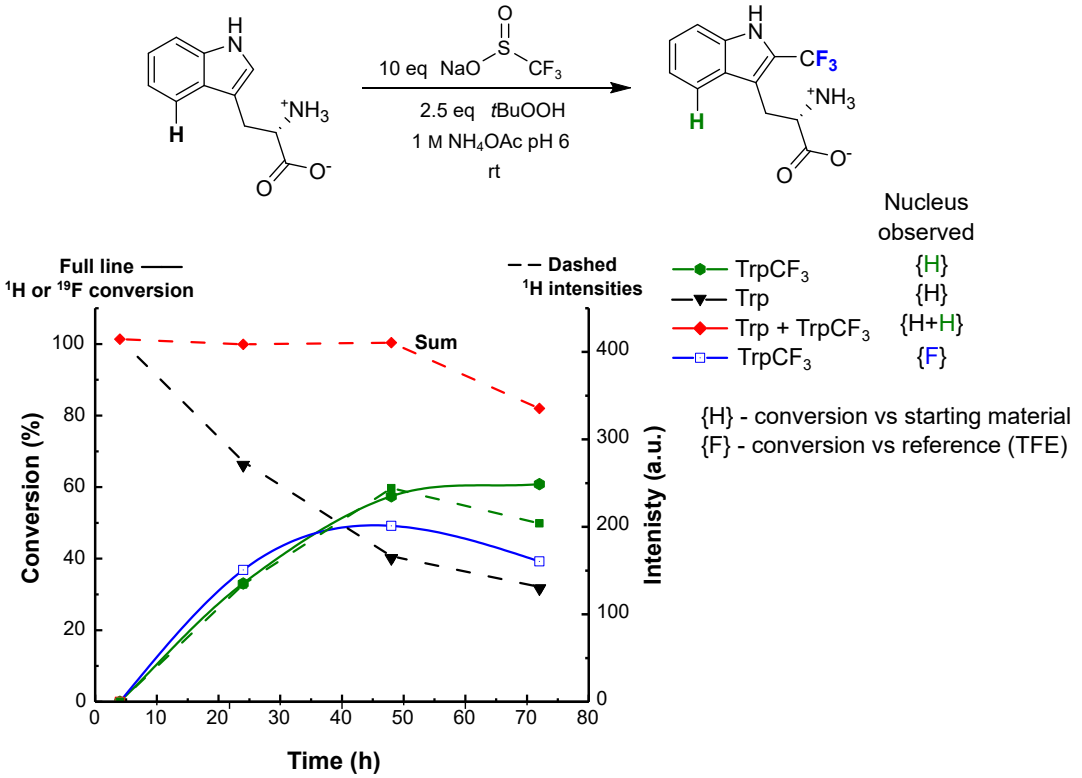
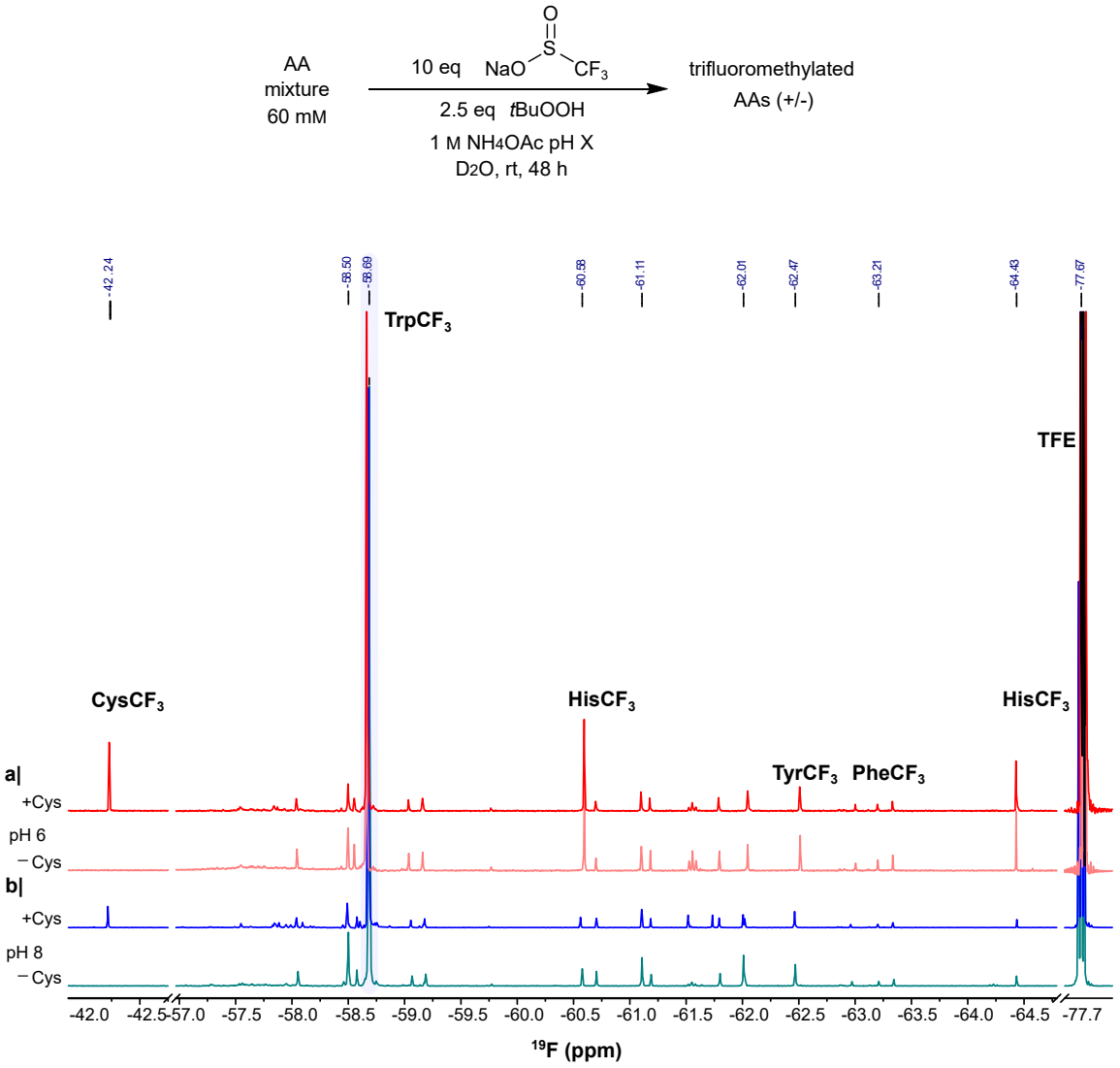


Fig. 2.4: Comparison of the ^{19}F and ^{1}H NMR monitoring of Trp trifluoromethylation. An analysis of the absolute intensities revealed the further reactivity of the **TrpCF$_3$** product.

This was aided by the additional quantitative ^{1}H NMR monitoring of the absolute **TrpCF$_3$** signal (C5 proton), with the conversion determined as a ratio of the product intensity versus combined substrate and product intensities. The experiment confirmed the clean formation of the desired product in the first phase of the reaction (<48 h), as ^{19}F and ^{1}H NMR conversions were in agreement and the sum of the ^{1}H NMR intensities for the product and the starting material remained constant. In the second phase of the reaction, both parameters started to vary, confirming the emergence of side products.

2.2.2 Selectivity

The next question was to determine if Trp residue chemoselectivity could be conserved in the presence of other reactive amino acid residues. To evaluate this, a simple limited conversion competition model was applied under various conditions. An equimolar mixture of the previously identified reactive amino acids (Trp, Tyr, Phe, His, Cys and Met as a potentially oxidisable additive) was prepared in a buffer of a desired pH. Each mixture was analysed at both pH 6 and pH 8 and with or without Cys. After 48 h, the crude reaction mixture was filtered and analysed by quantitative ^{19}F NMR to determine relative conversions - Fig. 2.5. This experiment revealed preferential formation of the desired product in all cases. The presence of free Cys, especially at high pH, reduced the conversion to the **TrpCF$_3$** product but had little influence on the relative selectivity. In its absence, at the higher pH, **TrpCF$_3$** was formed only marginally faster, with a $k_{\mathrm{rel}} \sim 5$ versus His. On the other hand, at lower pH, the desired product formation was substantially favoured, with a $k_{\mathrm{rel}} \sim 28$.



Amino Acid	Chemical shift	−Cys conv. (%)		+Cys conv. (%)	
	[ppm]	pH 8	pH 6	pH 8	pH 6
Trp	−58.7	28.8	36.8	19.5	33.8
His	−60.6, −64.4	4.1, 1.5	0.9, 0.4	2.6, 1.1	0.5, 0.3
Phe	−63.3	0.6	0.3	0.3	0.3
Tyr	−62.5	1.6	1.3	0.8	1.0
Cys	−42.2	-	-	1.5	0.9

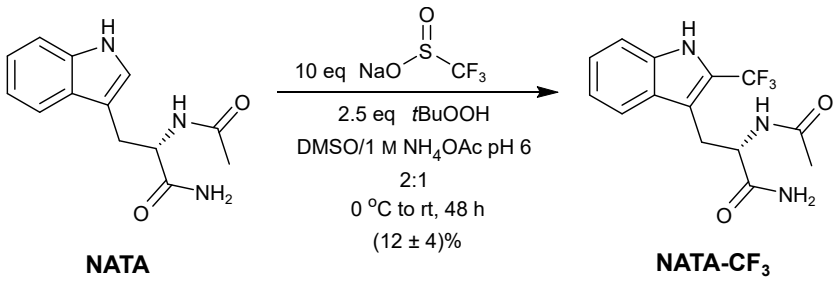
Fig. 2.5: Trifluoromethylation of Trp in direct competition with other reactive AAs. **a-b|** Fragments of the ^{19}F NMR spectra of the reaction mixtures in the competition experiment (with and without Cys) at different pH **a|** pH 8, **b|** pH 6. In the table: conversions to fluorinated products and the position of their signals.

The most likely explanation of this result is that the electrophilic trifluoromethyl radical has a lower affinity for the protonated His residues, present more abundantly at pH 6 ($pK_a \sim 6.6$)[239]. Even lower pH values were not considered as they are unlikely to be broadly applicable in protein chemistry.

Finally, the experiment was repeated in 10x lower concentrations (closer to concentration regimes for biomolecule modifications and ensuring the full solubility of Tyr), which yielded a similar result, with $k_{\text{rel}} \sim 15$ caused by a slightly higher Tyr reactivity. However, the overall **TrpCF$_3$** conversion was as low as 3%, indicating a very slow reaction in such conditions, which may not have reflection for protein modification.

2.2.3 Peptidic substrates

Having established chemoselectivity on Trp in the presence of other AAs, the reaction was assessed on peptidic substrates. Initially, the trifluoromethylation reaction was tested on the doubly protected Trp substrate NATA, which proceeded with the same regioselectivity, albeit in lower yield - Sch. 2.3. This was most likely due to the addition of DMSO as a co-solvent, needed for substrate solubilisation.



Sch. 2.3: Radical trifluoromethylation of *N*-acetyl-tryptophanamide (90 μmol, 45 mM, $n = 2$)

Next, melittin, a 26-residue peptide with a single Trp was chosen to further in-

vestigate the reaction. Importantly, this biomolecule does not have any other potentially reactive residues and the progress of the reaction can be monitored in a semi-quantitative way via LC-MS. It quickly became apparent that in order to achieve appreciable conversion using a low concentration of the peptide (∼hundreds μM), a much higher excess of the other reagents was needed - Fig. 2.6a. However, this also resulted in a rapid over-modification, the extent of which depended on the specific conditions used - Fig. 2.6b. Careful balance between the reaction time and the reagent excess had to be struck in order to observe only the singly trifluoromethylated product.

An optimal condition window was identified: short (∼mins) reaction time, $\sim (1 - 20)$ mM $\equiv$ hundreds of equivalents of the radical precursor and a magnitude lower excess of oxidant at temperature of 0 °C. LC-MS/MS analysis of the reacted peptide digested with trypsin revealed Trp19 as the only site of CF_3 modification - Fig. 2.6c. Analogously, oxidative damage side products, typically encountered for Trp residues[240], were also found, but in a lower relative intensity. It was also observed that the presence of Met was beneficial to reduce the undesired oxidation of the substrates. This was to be expected, as previous work has shown that Met serves as a sacrificial redox buffer[241]. Further optimisation of conditions was not performed as it was envisioned for the protein substrates of higher complexity.

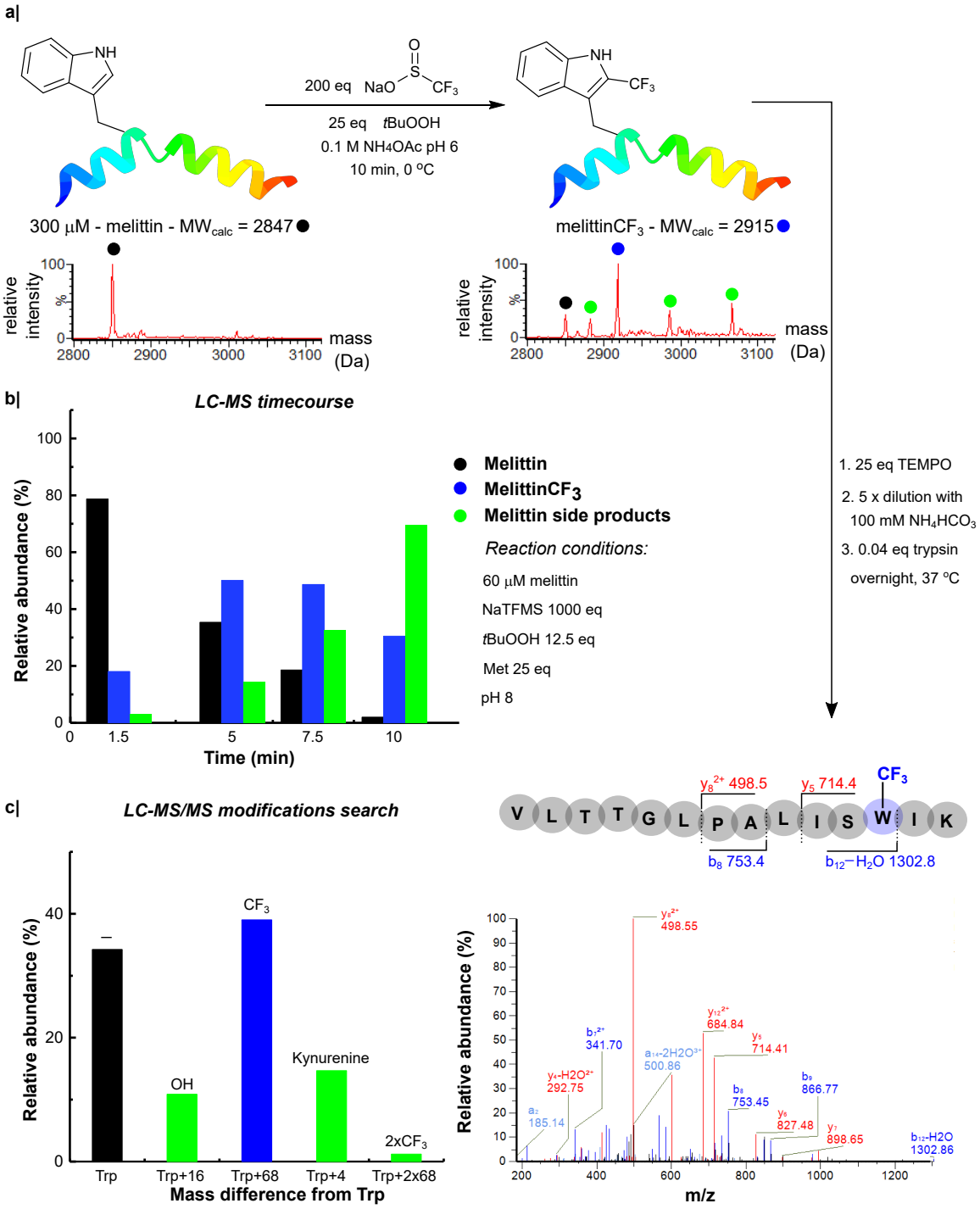


Fig. 2.6: Trifluoromethylation of the model peptide - melittin. **a**| Deconvoluted mass spectra for the starting material (black dot) and the reaction mixture after 10 mins, revealed conversion to dominant singly trifluoromethylated product (blue dot), among a low level of side products (green dots), **b**| LC-MS timecourse experiment under shown conditions, demonstrating the need for parameter screening for optimal conditions. **c**| LC-MS/MS analysis after tryptic digest identified Trp as the main site for both the desired (CF_3) and undesired modifications (with speculated identity based on the detected mass difference).

Together, the results from both the small molecule and peptide models suggested that upon optimisation of the reaction conditions, Trp selective trifluoromethylation may be possible for a more complex protein substrate.

2.3 Protein trifluoromethylation

2.3.1 Myoglobin as a model protein

Equine heart myoglobin (eqMb) was employed as a model protein, as it is an excellent representative example of globular proteins. This small, natural, and abundant protein possesses 153 amino acids (and a molecular weight of 16,951 Da), with 2 Trp (Trp7, Trp14), and 20 other aromatic residues (11 - His, 7 - Phe and 2 - Tyr) and importantly, no Cys. Interestingly, it was the first protein to have its structure precisely determined via X-ray crystallography[242]. This study revealed that the polypeptide chain is made up of 8 alpha helices, which in total make up 75% of the structure. The Trp residues are buried inside the compact globular structure, as are most of the other (hetero)aromatic residues. A theoretical investigation of residues solvent accessibility revealed that only His48, His81 and Tyr147 are relatively exposed - Tab. 2.2.

It is worth to keep in mind that the biochemical function of myoglobin is the storage of oxygen, enabled by a tightly but non-covalently bound prosthetic group: a ferrous heme that also causes this protein to be coloured and to absorb visible light additionally tying the protein structure with function[243].

Tab. 2.2: A solvent accessibility map of potentially reactive residues in myoglobin (PDB: 1YMB). Solvent accessible surface (SAS) was calculated with Chimera software (v. 1.11.2) using default probe radii[244]. The relative exposure (rel. exp.) was assessed by a comparison of the solvent excluded area by a Xaa residue in this protein versus one excluded in the tripeptide Gly-Xaa-Gly[245].

Residue	SAS [Å^2]	Rel. Exp. [%]	Residue	SAS [Å^2]	Rel. Exp.[%]
W7	15.4	29	H93	52.2	46
W14	6.6	35	H97	74.5	61
H24	7.8	22	Y103	36.7	42
F33	5.8	13	F106	28.2	37
H36	56.8	44	H113	81.5	67
F43	52.0	49	H116	68.1	57
F46	0.8	11	H119	33.2	39
H48	108.5	75	F123	8.8	30
H64	41.3	45	F138	20.5	49
H81	141.6	82	Y147	173.4	82
H82	5.0	23	F151	36.2	45

2.3.2 Reaction optimisation

The translation of the conditions developed for the peptide straightforwardly led to the clean trifluoromethylation of eqMb as evident by LC-MS analysis. To maximise the Trp trifluoromethylation (i.e. double trifluoromethylation, with no side products or starting material present), the reaction time and the reagent stoichiometry were investigated to find the optimal values. To ensure a general applicability (for more sensitive proteins), it was decided that Met as an additive as well as a lower temperature would constitute default conditions.

Similarly as for the model peptide, no products were observed using a low excess of reagents (NaTFMS, *t*BuOOH $<$ 50 eq). However, using higher amount of the reagents, a single trifluoromethylation modification occurred within minutes, with the

double addition product following shortly after - Fig. 2.7b-c. Therefore, 10 minutes was chosen as the standard reaction time. From there, it was found that an increase in the amount of the oxidant did not significantly increase the conversion - Fig. 2.7d, which was much more easily manipulated with an excess of the radical precursor - Fig. 2.7e. Again, using a much larger excess of NaTFMS (> 400 eq), the side products became the prominent species, highlighting the need of the timely termination of the reaction, here with the persistent radical - 4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl (TEMPOL), added just before LC-MS analysis.

The main side reaction observed was over CF_3-ylation with minimal appearance of oxidised species. This can be understood in light of the fact that sensitive residues are buried inside the hydrophobic core of the protein, and hence not readily accessible to the solvent (SAS for Met55: 1.56 Å^2, Met131: 0.64 Å^2). Conducting the reaction at lower pH 4, in which eqMb is partially unfolded[246], led to the appearance of an oxidation pattern, as evidenced by the characteristic $\Delta m = +16$ Da adducts in the mass spectrum. Interestingly, the protein was also slowly oxidised in neutral pH, but only in the absence of the radical precursor. However, since this reactivity was essentially absent if an excess of NaTFMS was added to the reaction mixture, and as such, full engagement of the oxidant in the radical generation process can be deduced.

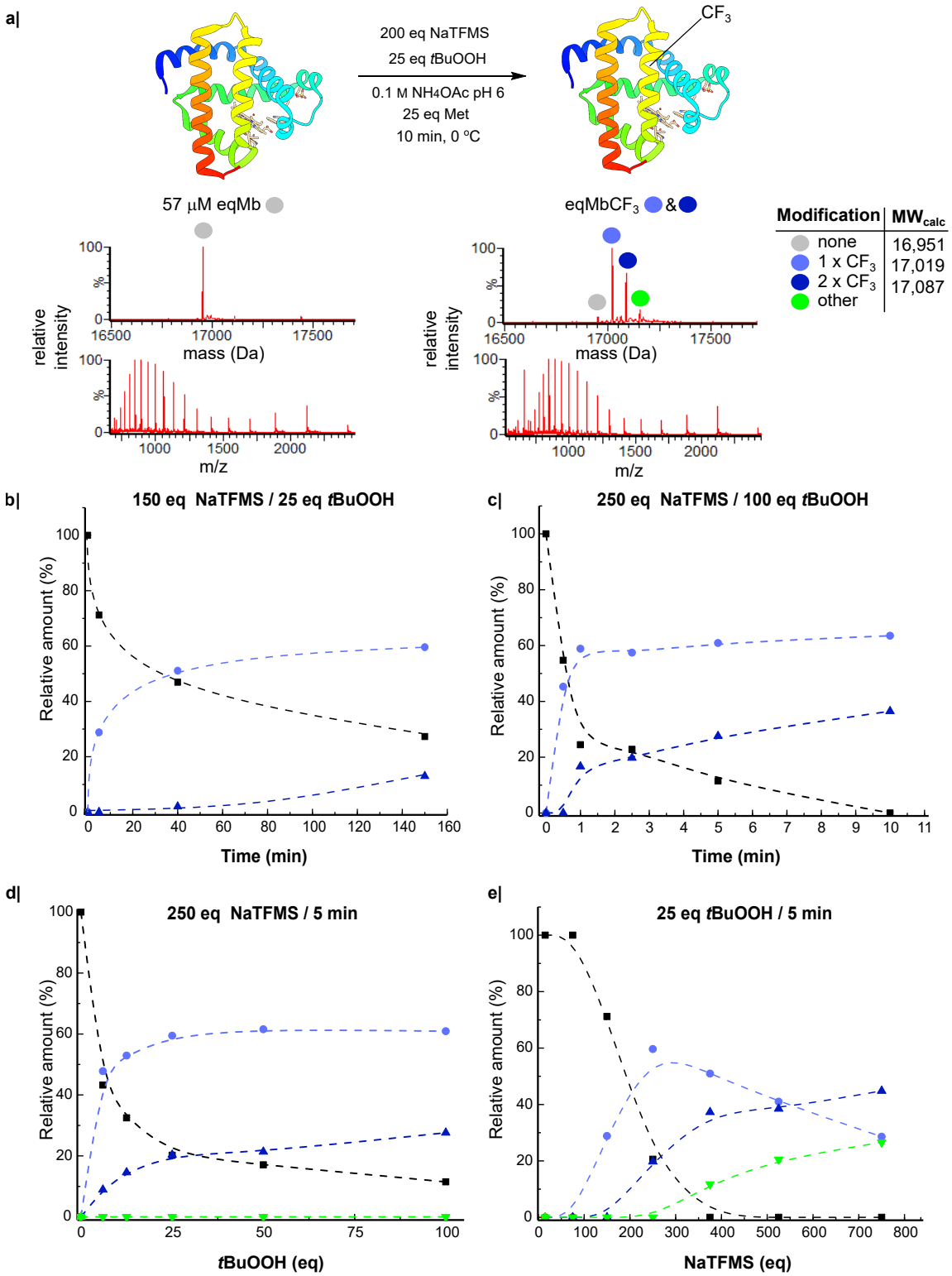


Fig. 2.7: Investigation of myoglobin trifluoromethylation. **a**| Reaction under optimised conditions with deconvoluted (upper) and unprocessed (bottom) ion series, revealing the formation of trifluoromethylated products (Δm = +68). **b-e**| Investigation of time and reagent excess influence on the formation of products.

2.3.3 Selectivity assessment

To confirm the modification site, a sample of modified eqMbCF$_3$ (under the optimised conditions) was digested with trypsin and the resultant peptides were analysed by LC-MS/MS - Fig. 2.8a.

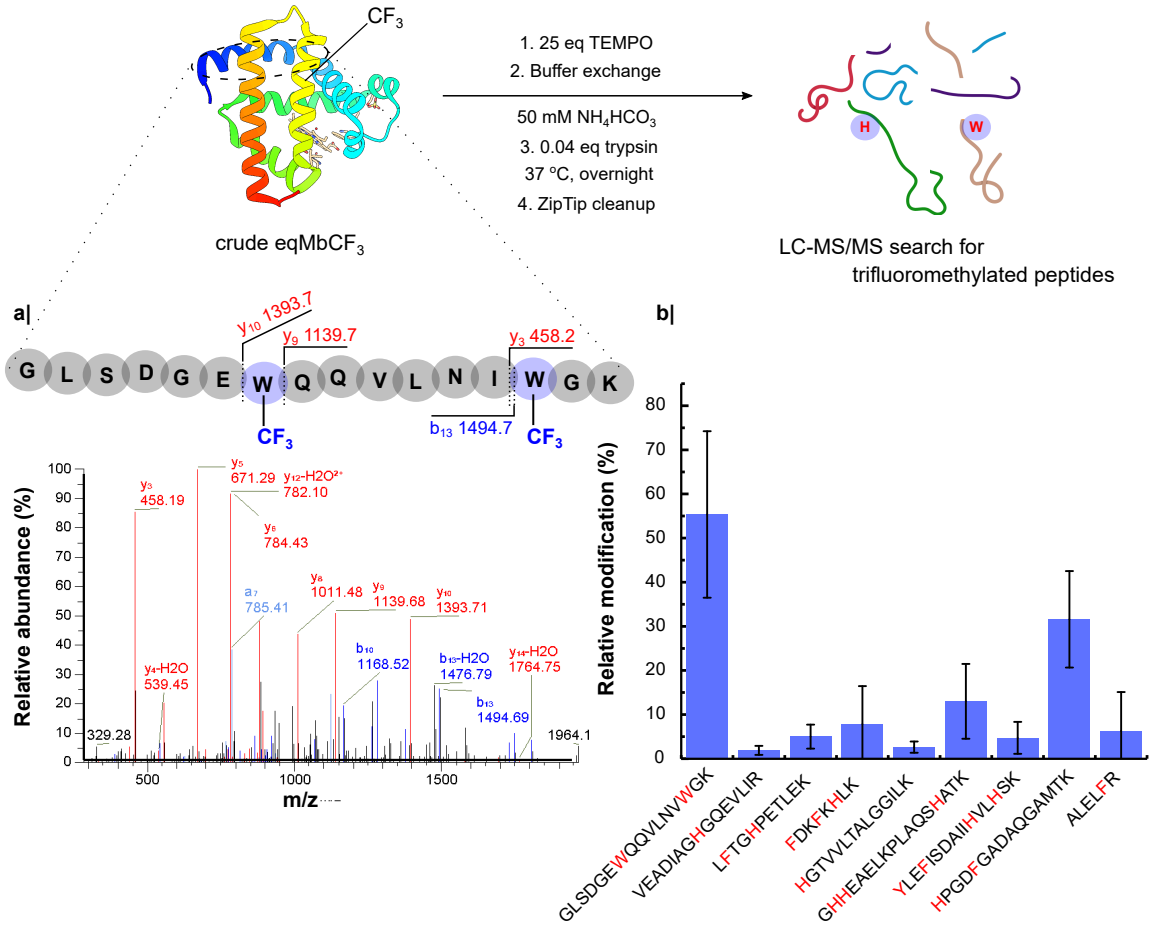


Fig. 2.8: Results of LC-MS/MS analysis ($n = 3$) of eqMbCF$_3$ reacted under standard conditions and digested with trypsin. **a**| The double CF$_3$-ylated peptide (at Trp residues) could be identified via its fragmentation pattern, **b**| Quantification of different tryptic peptides with single CF$_3$ group. The relative modification was calculated by an intensity comparison of the singly modified peptide (on one of the residues marked with red) to that of all forms of the peptide (unmodified etc.). Results were averaged for three samples, with the error bars indicating the standard deviation. The data was not normalised for the difference in the absolute intensity displayed by the peptides and therefore cannot be treated quantitatively.

Unsurprisingly, this revealed Trp7 and Trp14 as the main trifluoromethylated residues, among a low background of modified His - Fig. 2.8b. Due to the tryptic digest specificity a few shorter peptides were not detected. However, the overall coverage amounted to 86% and covered the key Trp residues as well as majority of the His residues (10/11).

In the next step, residual modification of AAs other than Trp was correlated with their calculated solvent accessible surface to verify if this could be a factor enhancing reactivity - Fig. 2.9a.

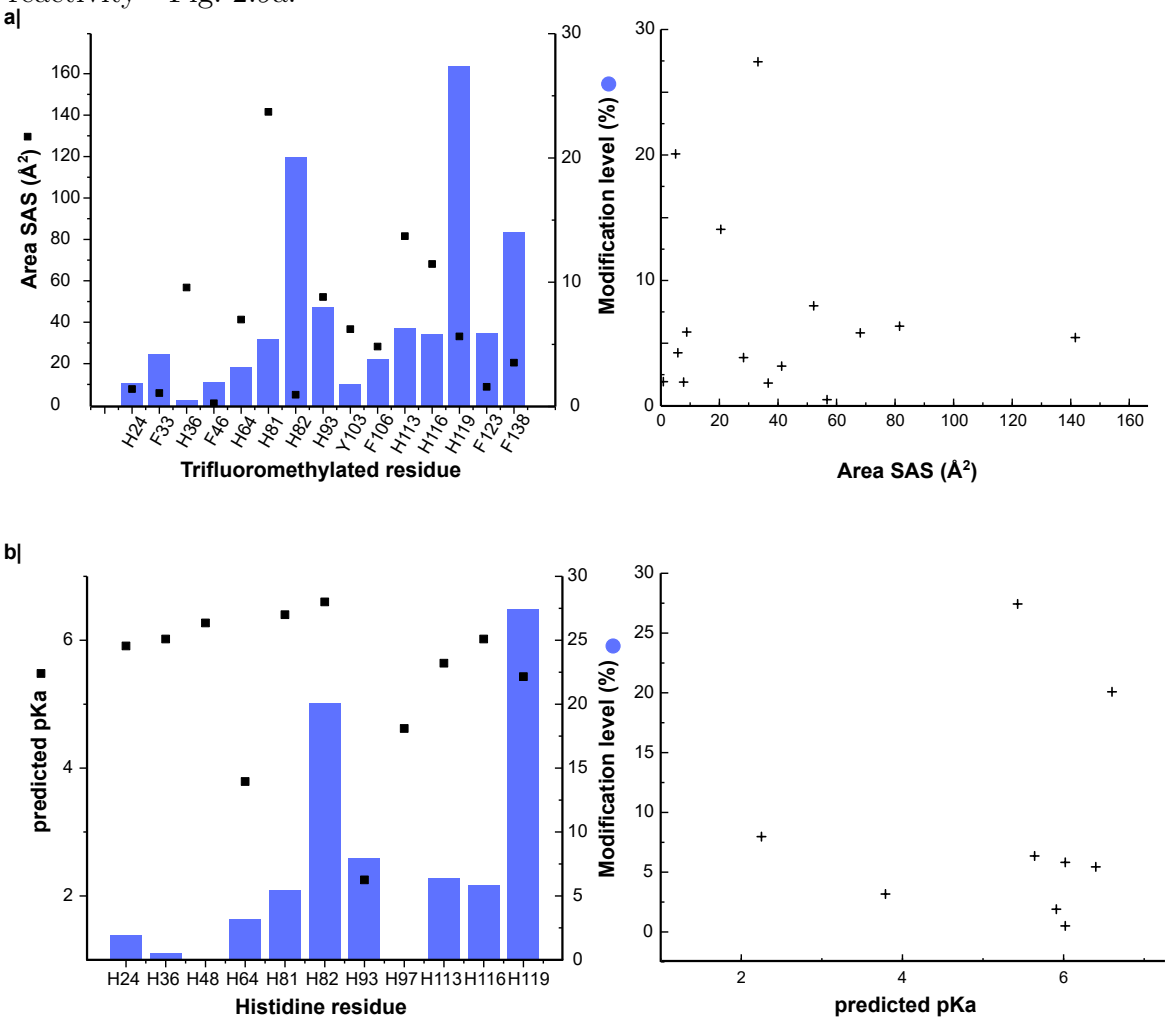


Fig. 2.9: Attempt to correlate observed conversions with certain physicochemical parameters: **a**| calculated solvent accessible surface, **b**| predicted pK_a value of His residue[247].

The Spearman's correlation coefficient equaled -0.01, suggesting that there is no correlation between the two variables. The same result was obtained when the analysis was based on the relative exposure (not shown). Similarly, no correlation between the reactivity of the His residues and their predicted pK_a was found - Fig. 2.9b, with a Spearman's coefficient of -0.13. These results suggested that the modification might be statistically distributed across the residues or that the MS measurements did not yield accurate quantitative data preventing this type of analysis.

In the next step, NMR was employed to confirm the selectivity of the trifluoromethylation procedure. To prepare a sample, the reaction was upscaled 200 times, which allowed recovery of >70% protein was post modification as quantified via UV-VIS spectroscopy. The spectrum of natively folded eqMbCF_3 displayed two main peaks in the region of interest, most probably arising from two different Trp residues in non-equivalent positions within the folded protein - (Fig. 2.10a - blue). To prove this, the protein was either denatured with 10 M urea - (Fig. 2.10a - grey) or digested with trypsin - (Fig. 2.10a - black) - both of the operations should led to normalisation of the local environment and an effective collapse of the chemical shift to this displayed by the small molecule **TrpCF_3** - (Fig. 2.10a - shaded, -58.7 ppm). This was indeed observed in the experiment.

Moreover, the superior Trp selectivity was confirmed as, although visible in the spectrum, peaks corresponding to modified His accounted for only ∼20% of the total signal intensity. From this result, considering 5.5 fold excess of the His residues and assuming no protein environment effect, the reaction with Trp was approximately >20 times faster ($k_{rel} \sim 24$) than that with His, and is in rough agreement with the

conclusions from the model studies.

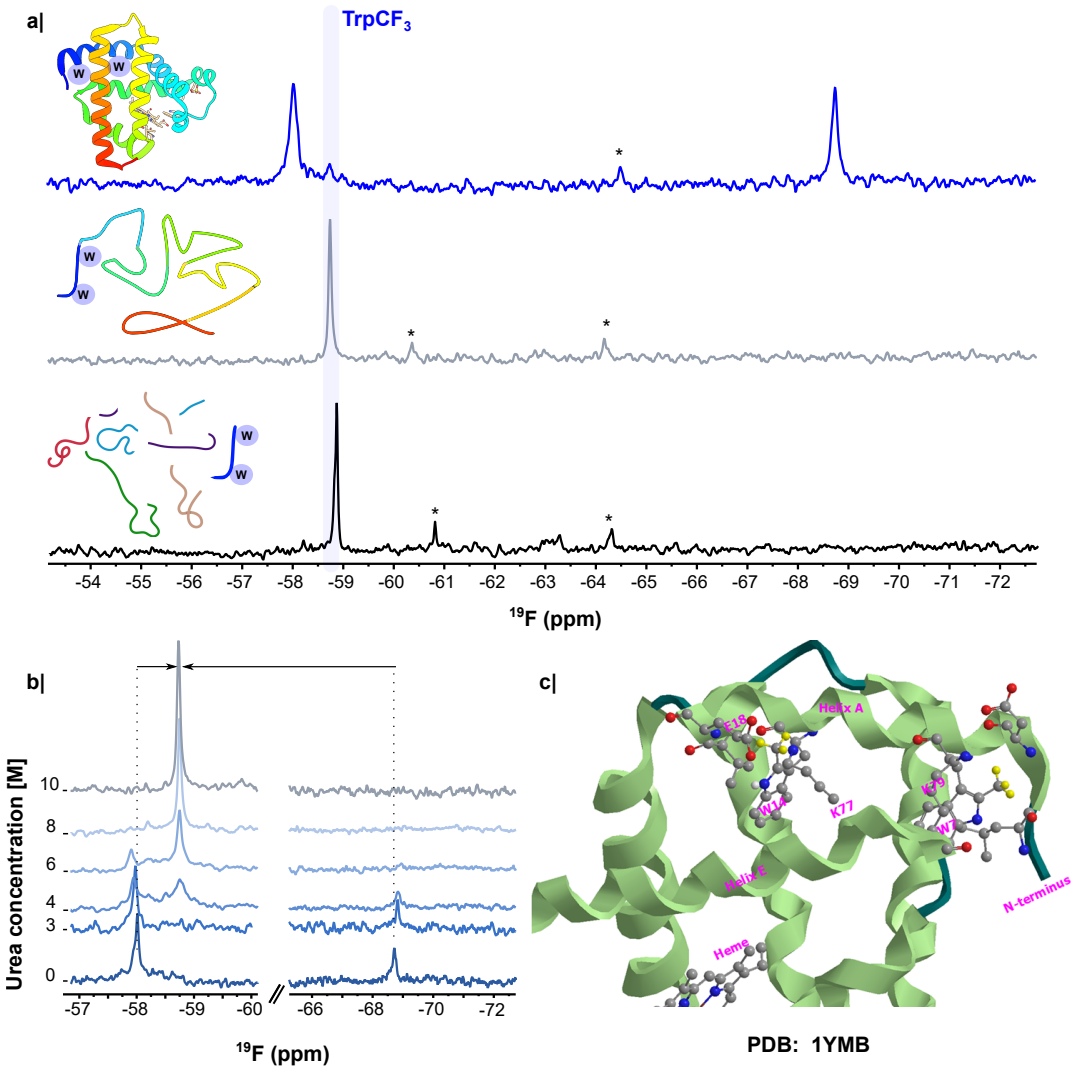


Fig. 2.10: **a**| Fragments of ^{19}F NMR spectra (processed with a 15 Hz line broadening) of eqMbCF$_3$ samples under different conditions: **a**| native (blue - 700 μM), denatured by 10 M urea (grey - 300 μM) and after tryptic digest (black - 2 mM). In all cases 0.1 mM NH_4OAc, pH 8 was used as the buffer. The chemical shift for 'freed' Trp residues closely matched that of the small molecule model (shaded area). Note the appearance of peaks (with asterisk) at -60.4 ppm and -64.2 ppm which may correspond to the chemical shifts of modified His. **b**| Urea titration experiment revealed a gradual denaturation of eqMbCF$_3$, suggesting co-existence of both denatured and folded forms at intermediate concentrations, **c**| Tentative visualisation of eqMbCF$_3$ (PDB structure: 1YMB, rendered with Chem3D 15.1 software, PerkinElmer Informatics) with the putative location of fluorine atoms in the CF_3 group marked in yellow. The heme and atoms within 6 Å from Trp residues are shown.

In order to make a tentative assignment of the two **TrpCF$_3$** peaks, a following rationale was based on their distance to the 'original' chemical shift:

(a) Trp14 is considered to be more buried, as supported by calculations in Tab. 2.2, and therefore makes more hydrophobic contacts. As such, it has a significantly different micro-environment when compared with a small molecule Trp that freely interacts with the solvent.

(b) Trp14 was reported to be transiently locked by a hydrogen bond to Glu18, which is likely to cause a significant chemical shift perturbation of the nearby fluorine atoms[248].

(c) a parallel comparison with a universally labelled with [4F]Trp Mb (in which all Trp residues contains fluorine instead a hydrogen at C4 position) reveals a similar peak dispersion; the assignment was achieved by a selective mutagenesis experiment[249].

Based on these assumptions, the peak at -58.0 ppm was identified as Trp7, while the one at -68.7 ppm as Trp14. The modified residues are tentatively visualised in Fig. 2.10c.

Interestingly, the full ^{19}F NMR spectra for the denatured or digested protein revealed the presence of new, low intensity peaks, in an unexpected region (~ -76 ppm) - Fig. 2.11. To test the hypothesis of the heme trifluoromethylation and side products formation due to the peroxidase activity of myoglobin[250], the modification was performed for apomyoglobin (ApoEqMb), in which this prosthetic was first extracted

with an organic solvent. Essentially the same results were obtained with additional peaks appearing upon the digestion, indicating no participation of the heme in the trifluoromethylation reaction. This was further confirmed via observation of the unmodified mass (616 Da)[251] during usual LC-MS analysis. This leave a hypothesis of side products small molecules possibly bound to myoglobin. Not interfering with the main conclusions, the issue was not further investigated, as additionally, the formation of these peaks were not observed during trifluoromethylation of different proteins.

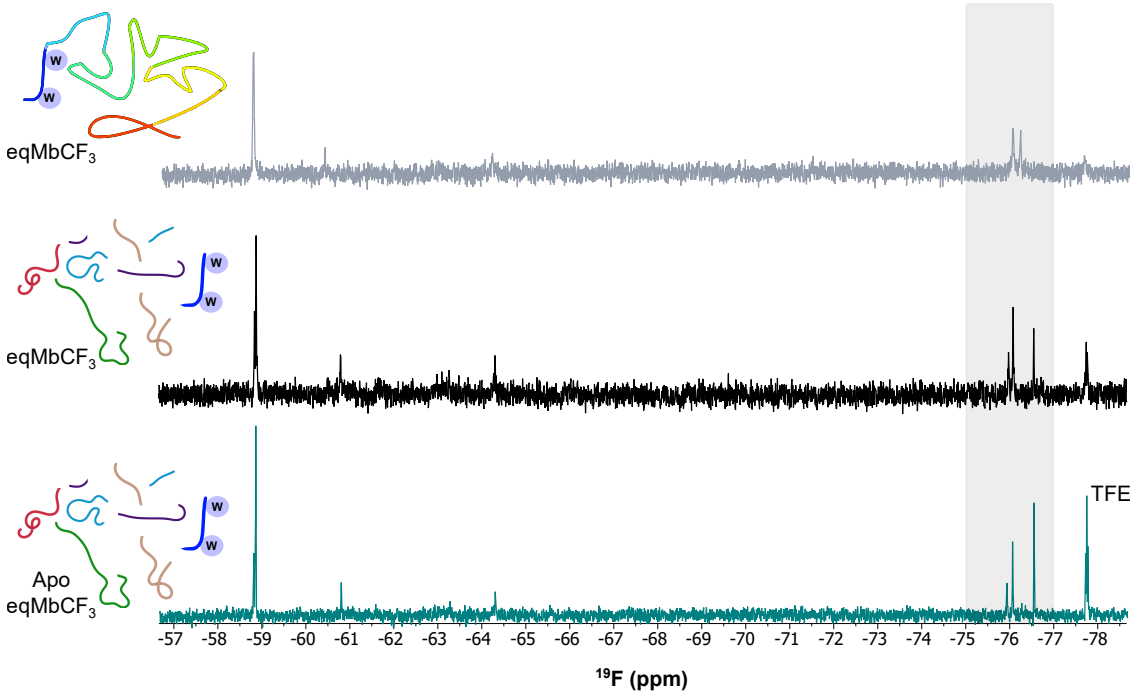


Fig. 2.11: The full ^{19}F NMR spectra (processed with a line broadening of 1 Hz) of trifluoromethylated proteins samples in denaturing assays. eqMbCF$_3$ was denatured with 10 M urea (grey - 300 μM) or digested with trypsin (black - 2 mM) while ApoEqMbCF$_3$ was digested with trypsin (teal - 2 mM). These conditions revealed unknown peaks highlighted in the shaded region, absent from the spectrum of a folded protein.

2.3.4 Characterisation of eqMbCF$_3$

Having determined the selectivity of the CF$_3$-ylation reaction, the focus was next turned to the biophysical characterisation of the model eqMbCF$_3$. In particular, an assessment of the protein structural integrity was critical, considering the fact that Trp modification occurred for the buried residues in the native state of the protein. To determine the effect of trifluoromethylation on the secondary structure, circular dichroism (CD) spectra were measured for samples with varying degree of the modification - Fig. 2.12a.

Pleasingly, the overlaying curves indicated that trifluoromethylation did not induce significant changes to the basic protein fold, confirming the lack of structural perturbations due to the minimal size of the CF$_3$ label. To further evaluate the impact of the modification on the stability of the model protein, CD spectroscopy was used to record melting curves and hence to investigate the thermal unfolding of the proteins - Fig. 2.12b. The midpoint of this transition, which corresponds to the melting point (T_m), was similar for all of the samples indicating comparable thermal stability Fig. 2.12c. The melting point for the unmodified protein was in agreement with values reported in the literature: (77 - 84) °C, obtained using the same method. Importantly, these results also suggested that the CF$_3$ modification did not affect the binding of the protein to the heme co-factor, as ApoEqMb was demonstrated to display a significantly lower stability[252,253].

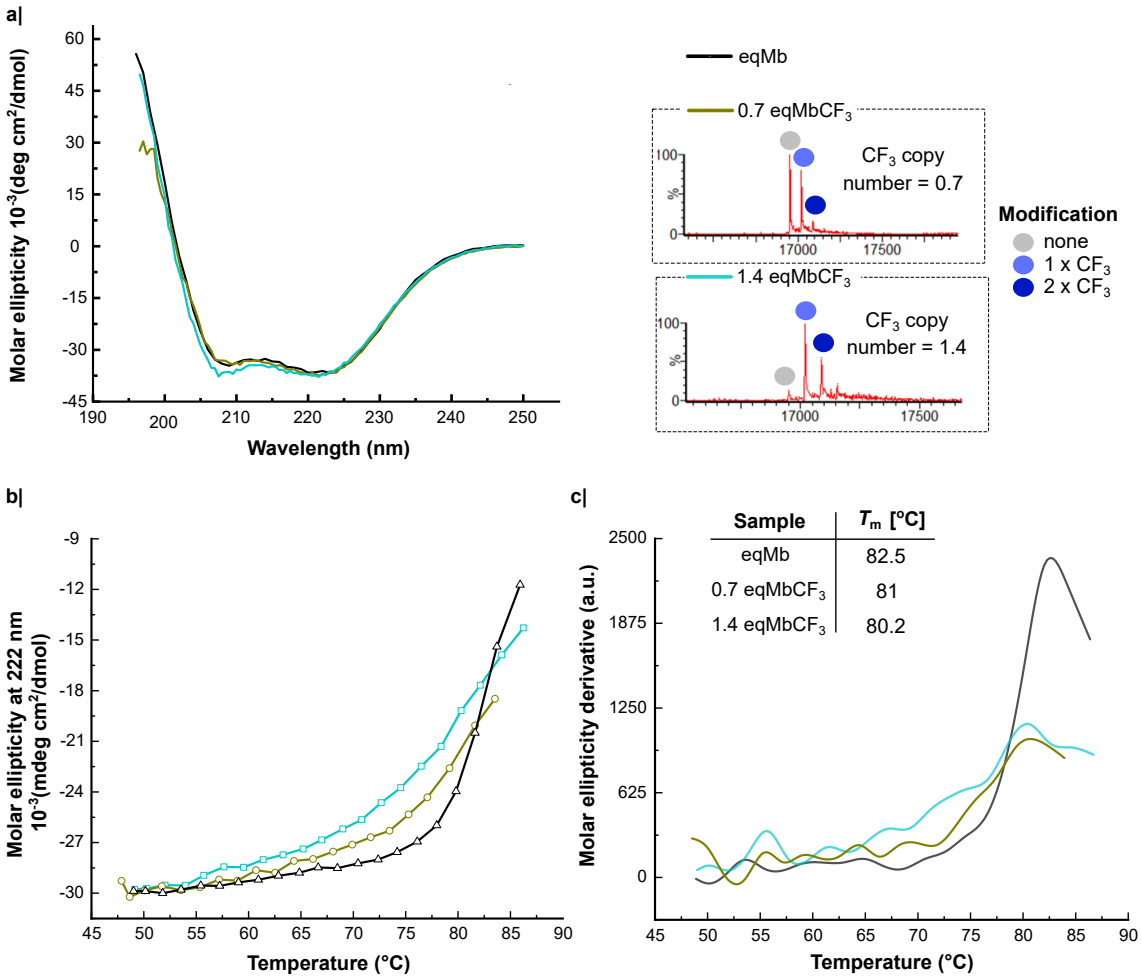


Fig. 2.12: **a**| Comparative CD spectra for an unmodified eqMb and two samples of eqMbCF$_3$ modified to a different degree (by the variation of the reaction time, ∼0.2 mg/mL protein in 10 mM PBS pH 8.0). LC-MS characterisation revealed the average CF_3 copy-number for different samples. **b**| Molar ellipticity at 222 nm, corresponding to the α-helical content of the protein was measured as a function of the temperature to evaluate thermal stability of the samples **c**| The first derivative of the molar ellipticity at 222 nm was used to determine melting temperatures (T_m).

Considering the fact that eqMb displays unique optical properties, UV-VIS spectroscopy was used to confirm the presence of heme and an assessment of its ligation state - Fig. 2.13a. The maxima positions for the Soret band and longer wavelength shoulders remained unchanged, further indicating that the trifluoromethylated sample

retained the heme with its native ligand (water molecule)[254].

Finally, the integrity of the protein was assessed by SDS-PAGE analysis, which displayed a single protein band for eqMbCF$_3$, regardless of the degree of the modification - Fig. 2.13b.

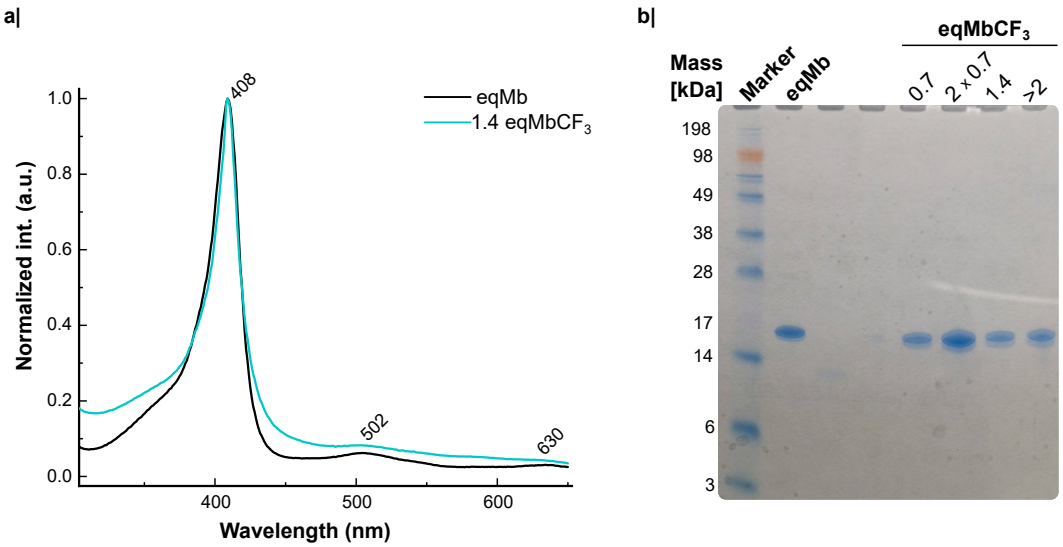


Fig. 2.13: **a**| A comparison of normalised UV-VIS spectra of eqMb and eqMbCF$_3$ ($\sim$50 μM in NH_4OAc pH 6.0, CF_3 copy number $\sim$ 1.4). **b**| An SDS-PAGE analysis of the eqMbCF$_3$ with different levels of CF_3 copy number ($\sim$ 10 μg of a protein per lane, '2 x 0.7' label indicate a double loading of the 0.7 eqMbCF$_3$ sample).

2.3.5 Expansion of the substrate scope

To assess if the trifluoromethylation protocol can be used universally, the reaction was carried out on several additional proteins with different Trp contents and ratios of other competitive residues, presented in Tab. 2.3.

Tab. 2.3: A list of proteins used in the assessment of the developed trifluoromethylation protocol. The data for the commercially available proteins was compiled using UNIPROT database (eqMb = MYG_HORSE, Tr = TRYP_PIG, Ly = LYSC_CHICK), and for others, the sequence was known from the plasmid construction.

Abbrev.	**Name**	**Size [kDa]**	**#Trp**	**#His**	**#Tyr**	**#Phe**
A5	annexin V	35.8	1	3	12	13
PanC	pantothenate synthetase	34.1	1	12	7	9
SRP19	signal recognition particle 19 kDa	18.2	1	7	5	3
eqMb	myoglobin	16.9	2	11	2	7
Tr	trypsin	23.4	4	4	8	4
Ly	lysozyme	14.3	6	1	3	3

The proteins were reacted under standard conditions (see Fig. 2.7), and, pleasingly, in all cases trifluoromethylation products were observed via LC-MS. Specifically using that protocol, proteins containing only 1 Trp residue (A5, PanC and SRP19) still contained $>$ 50% of unreacted started material, therefore in order to improve the conversion the reaction was further optimised for these substrates. It was found that a small adjustment of the amount of radical precursor (+100 eq NaTFMS), additive (+25 eq Met) and the reaction time (5 mins) generally produced a greater amount of the desired product i.e. dominant species with one CF_3 group added - Fig. 2.14. Regrettably, it was impossible to prevent the generation of a small amount of the doubly modified protein, which presumably arised due to functionalisation of one of the other reactive residues (e.g. His or Tyr).

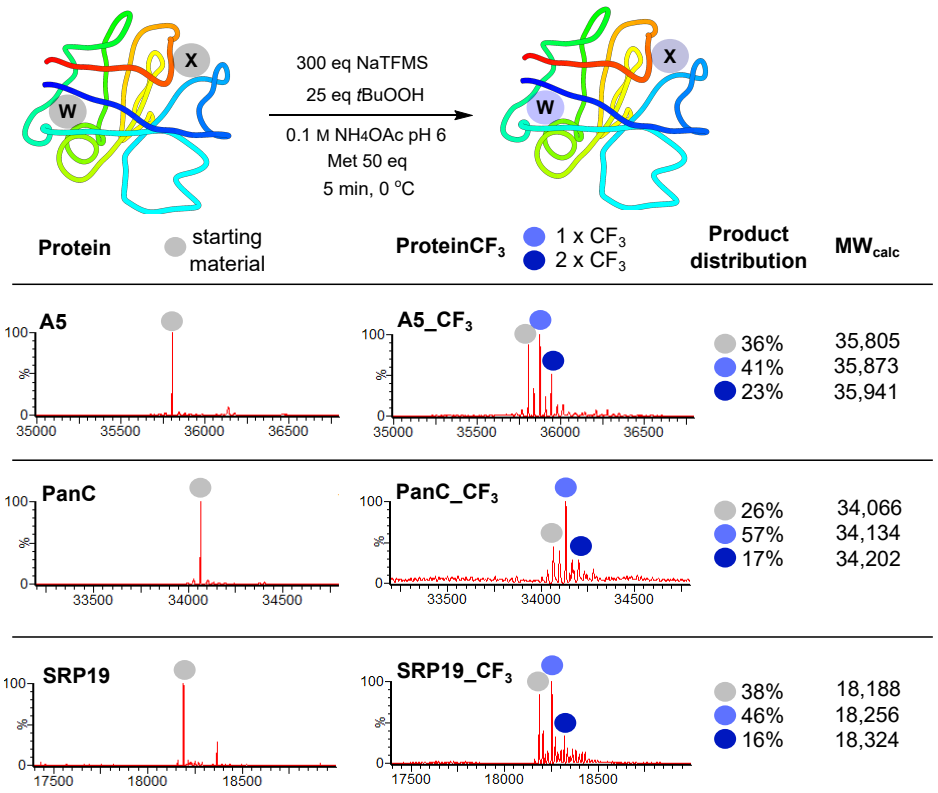


Fig. 2.14: Trifluoromethylation of proteins ($\sim (50-300)\ \mu$M) with a single Trp residue. LC-MS analyses revealed conversions of $> 60\%$ of the material to the fluorinated products, with predominant formation of a single CF_3 addition species.

Protein substrates chosen for the further investigation of the reaction were characterised by a different ratio of reactive residues to Trp residues. To gain further insights into the relative selectivity, comprehensive ^{19}F NMR analyses were carried out - Fig. 2.15. Importantly, from the synthetic point of view, all proteins were recovered in $(65-80)\%$ after the sequence of modification and buffer exchange.

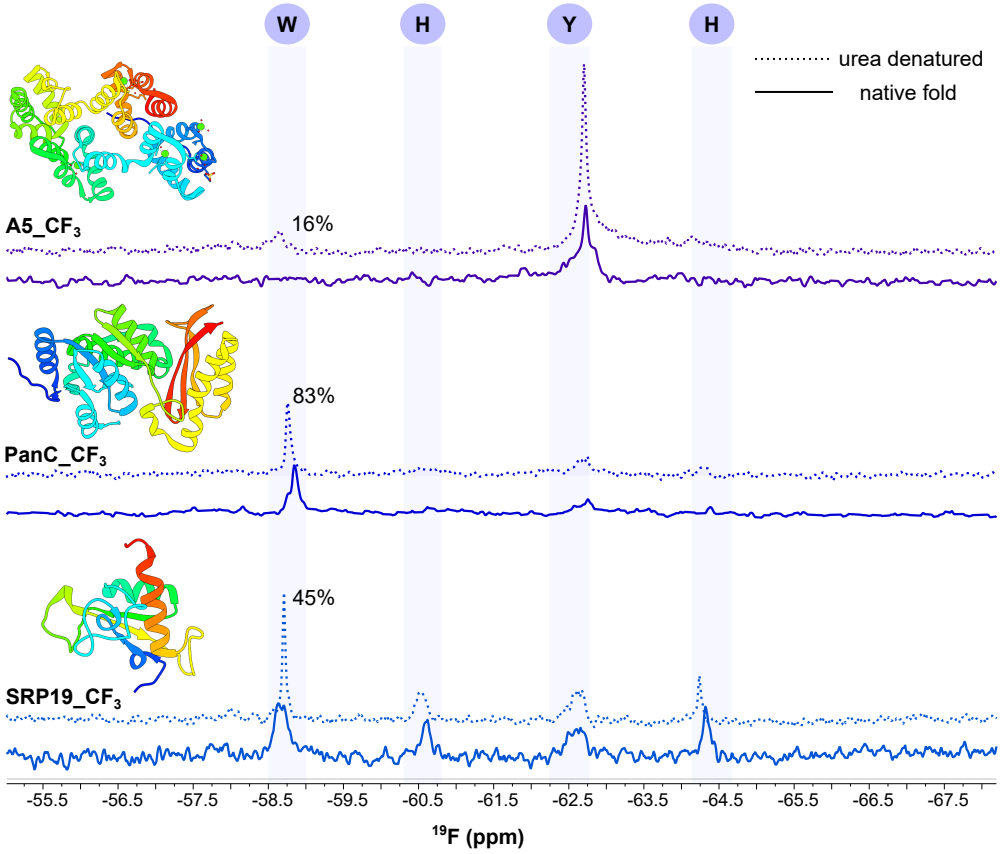


Fig. 2.15: Fragments of ^{19}F NMR spectra of trifluoromethylated proteins with a single Trp residue ($\sim (50-250)$ μM). Solid lines show spectra for samples in the native state (folding buffer 100 mM NH_4OAc pH 6), while dotted lines show spectra for the denatured samples (with 6 M urea). The peaks in the highlighted regions corresponding to trifluoromethylation of specific residues were integrated in order to evaluate the relative distribution of the CF_3 modification for Trp.

First, spectra under native folded conditions revealed the presence of a single peak in cases of PanC and A5, while SRP19 displayed four distinct signals. To elucidate their origin, all proteins were denatured by addition of urea, which sharpened the peaks (raising the prospects of quantitative precision of the analysis) and ensured their migration to the characteristic chemical shift regions. Integration of the peaks revealed that good Trp selectivity was observed only for PanC (83%) and partially for SRP19 (45%), while the main peak in the A5 spectrum corresponded to modified Tyr

residues (84%). This somewhat surprising result can be understood by appreciating the extreme 12-fold excess of Tyr residues over Trp. Importantly, this result suggested that for practical purposes, trifluoromethylation of proteins with high Tyr versus Trp content might not be Trp selective. One should note that for proteins with a similar His to Trp ratio (12 to 1 as in PanC) the reaction remained predominantly Trp selective suggesting an overall higher Tyr cross-reactivity. The high selectivity in the case of PanC was additionally verified via an LC-MS/MS analysis of the trypsin hydrolysate - Appendix (Fig. 2.29).

Next, in order to further understand the chemoselectivity and regioselectivity of the transformation, proteins with more than one Trp residue were evaluated. As expected, trypsin gave trifluoromethylated products under standard conditions with an average CF_3 copy number of 0.9 - Fig. 2.16a. As in the previous experiments, NMR analysis was performed to elucidate the modification site. In the case of Tr, a relatively complex ^{19}F NMR spectrum for the folded protein most likely indicated a non-trivial trifluoromethylation pattern in addition to conformational effects - Fig. 2.16b. Simple denaturation with a chaotropic agent (urea, guanidine) was not sufficient to fully unfold this protein, as no changes in the spectrum was observed. Only after an exhaustive denaturation with 8 M urea, 250 mM DTT and 10 min incubation at 85 °C, a simpler spectrum could be obtained with the main peak corresponding to Trp modification (59%). A prominent peak in the region of Tyr modification suggested that additional factors modulating the reactivity of specific residues may be important when considering the trifluoromethylation selectivity.

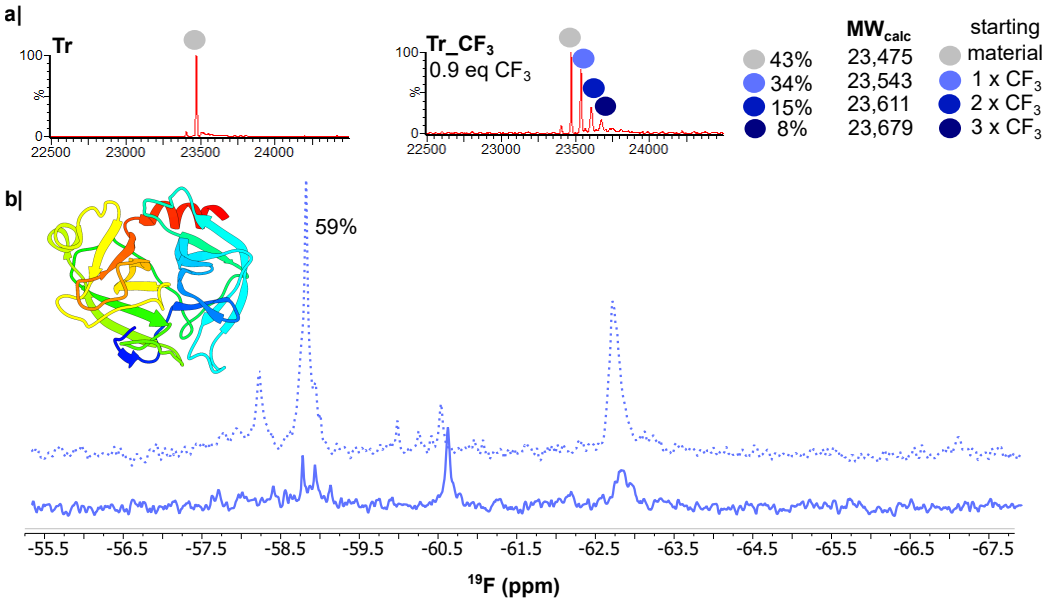


Fig. 2.16: **a**| Mass spectra for trifluoromethylation of Tr under standard conditions revealed an average addition of 0.9 CF_3 groups. **b**| Fragments of the ^{19}F NMR spectra of Tr_CF_3 ($\sim$ 200 μM) in the native state (solid line, folding buffer 100 mM NH_4OAc pH 8 + 10 mM $CaCl_2$) and denatured (dotted line, previous buffer and 8 M urea, 250 mM DTT after a 15 min incubation at 85 °C).

The next tested protein with multiple Trp was lysozyme (6 Trp residues), which under standard trifluoromethylation conditions yielded on average modification with 2.2 trifluoromethyl groups - Fig. 2.17a. In the native ^{19}F spectrum, broad and overlapping peaks were observed - Fig. 2.17b. Interestingly, the spectrum did not undergo full simplification when the protein was subjected to the same denaturing conditions (8 M urea, 250 mM DTT and 15 min incubation at 85 °C).

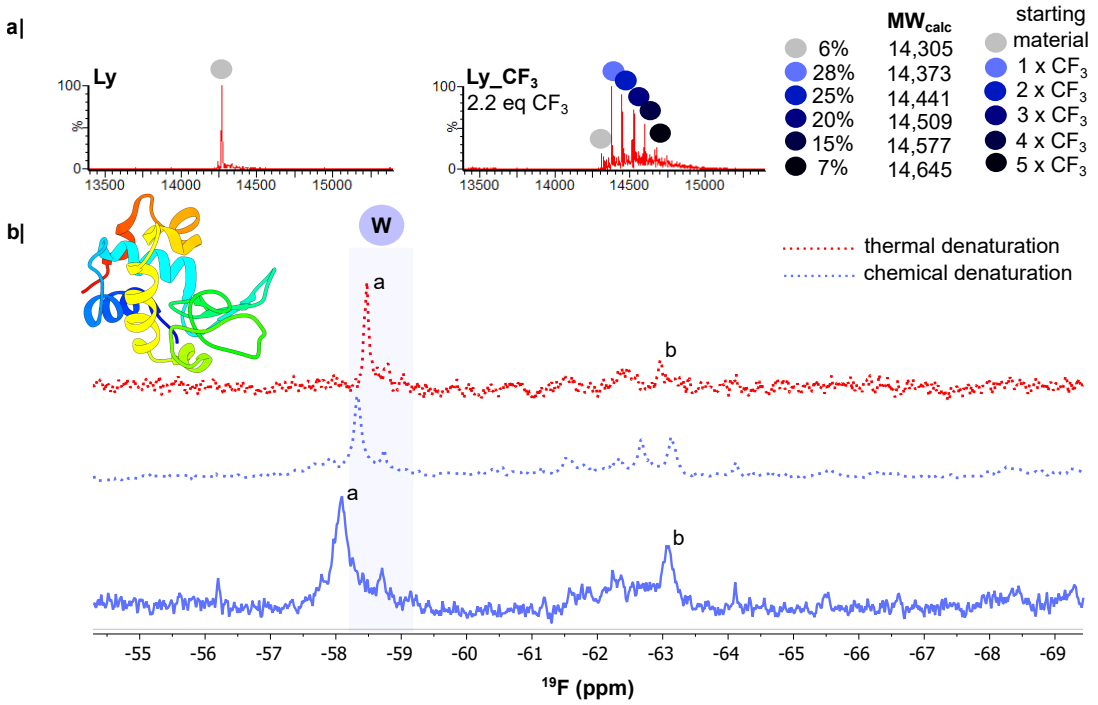


Fig. 2.17: **a**| Mass spectra for Ly trifluoromethylated under standard conditions revealed an average addition of 2.2 CF_3 groups. **b**| Fragments of ^{19}F NMR spectra of Ly_CF_3 ($\sim$ 400 μM) in the native state (solid line, folding buffer 100 mM NH_4OAc pH 6), chemically denatured (blue dotted line, previous buffer and 8 M urea, 10 mM DTT after a 10 min incubation at 85 °C) and thermally denatured (red dotted line, previous buffer, sample at 70 °C while measured).

Instead, thermal denaturation turned out to yield observable effects in the expected peak migration, while still not shifting to the peak locations associated with random conformation before the sample precipitated at its thermal limit [255] - Fig. 2.18a. These analyses helped to establish that trifluoromethylation indeed occurred predominantly for Trp residues. The increasing intensity of the peak corresponding to **TrpCF$_3$** (peak a), while decreasing of the others - Fig. 2.18b, indicated that the peak at -63 ppm (peak b) might also originate from Trp modification. Overall, the result for thermally denatured Ly_CF_3 showed high Trp selectivity, although the broadness of the other background peaks made it hard to give a reliable estimate.

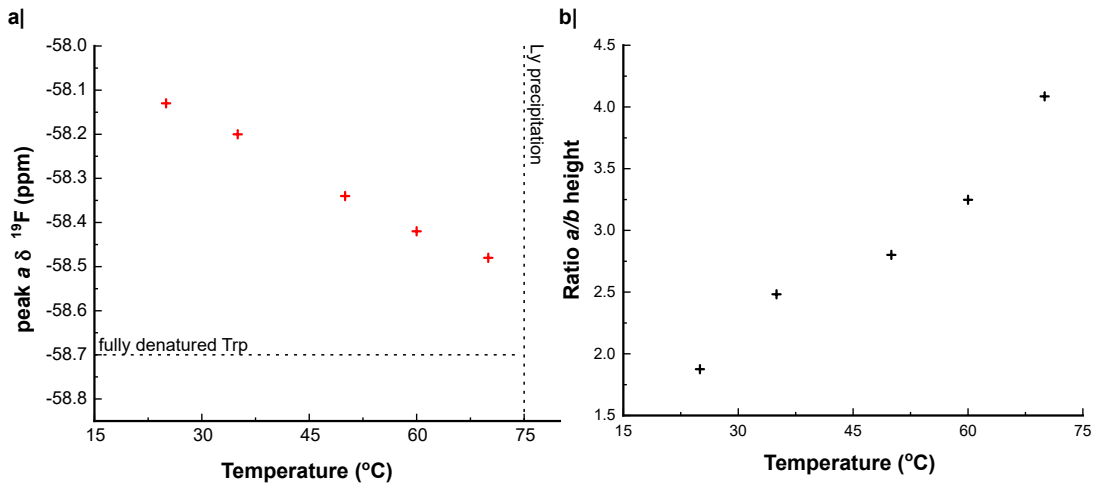


Fig. 2.18: Analyses of peaks position and intensity during the thermal denaturation of Ly_CF_3. **a**| Migration of Trp peak (peak a) as a function of temperature. The sample precipitated at the expected temperature[255] of 75 °C, before the peak reached the chemical shift of small molecule **TrpCF$_3$**. **b**| Estimation of a shift to the expected signal by a calculation of height ratios between peak a and b as a function of temperature. Gradual denaturation of the protein progressively caused an increase of the Trp signal at the expense of peak b.

In order to better understand the distribution of the trifluoromethylation, samples of Ly_CF_3 with a different level of modification were digested with trypsin and the resultant peptides analysed with LC-MS/MS. To compare the results semi-quantitatively, a relative conversion to single trifluoromethylated peptides was calculated - Fig. 2.19. The results did not indicate a single major target, with the bulk of the modifications found on Trp residues. The highest conversion was obtained for the tryptic peptide WWCNDGR with a double Trp motif containing the most solvent accessible Trp residue - Trp62. On the other hand, the residue with the lowest accessibility, Trp28, displayed the lowest conversion, as expected. However, the lack of significant differences between extensively and minimally modified samples (indicating a possible bias of the methodology), as well as between the other Trp residues, suggested a more

statistical distribution of the trifluoromethylation with a very low steric factor. This analysis revealed that five out of six Trp residues were found to be trifluoromethylated.

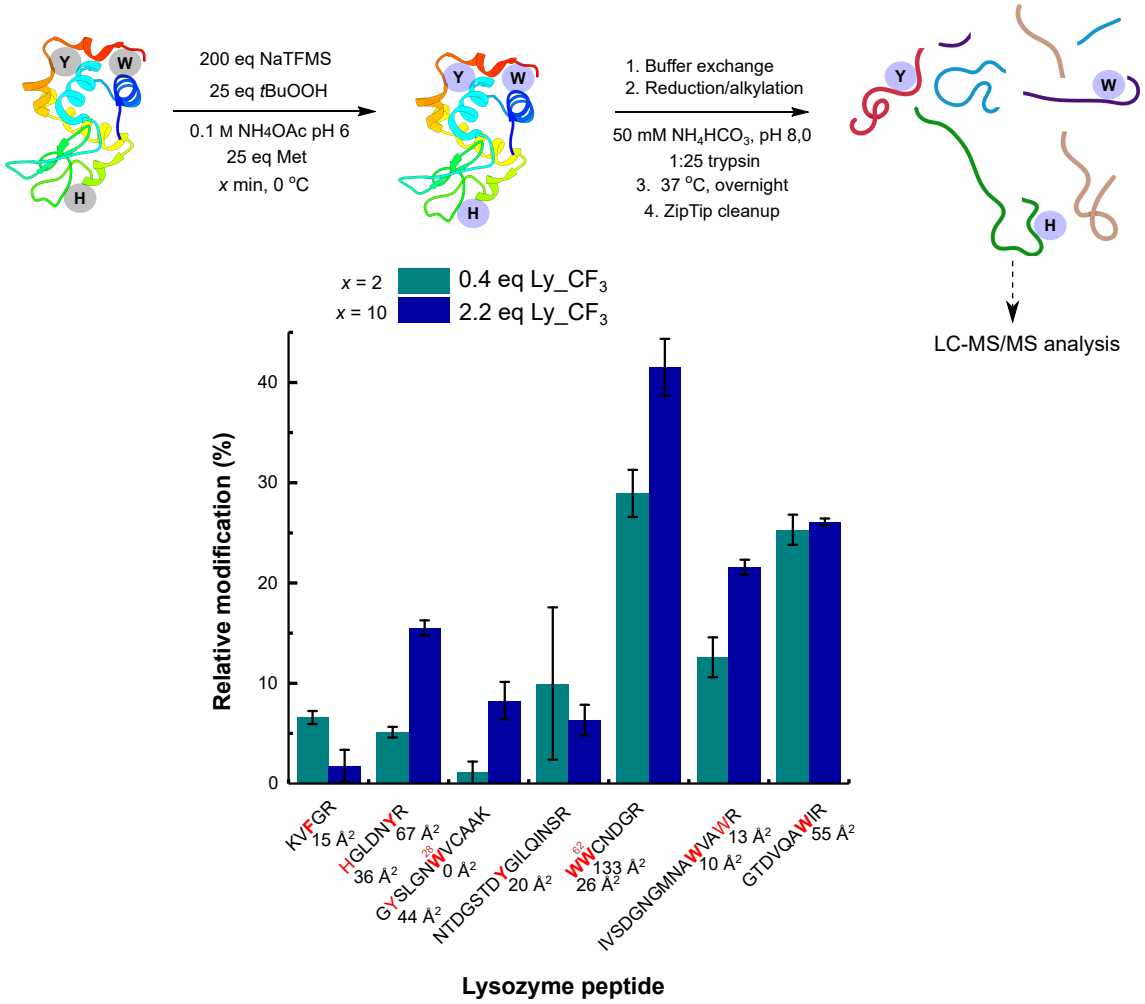
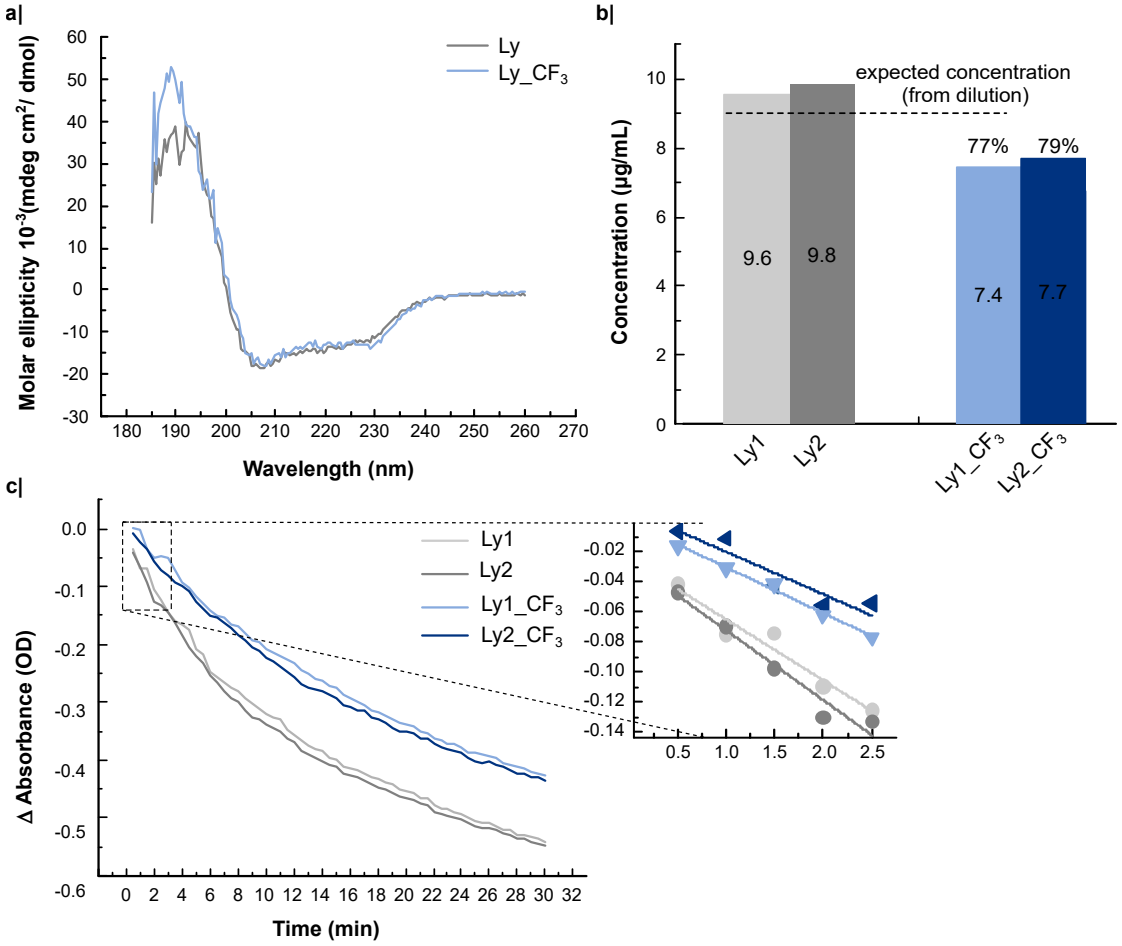


Fig. 2.19: Results of LC-MS/MS analysis ($n = 3$) for Ly_CF_3 samples modified to a different degree (with a 0.4 or 2.2 CF_3 copy number) showing a relative single CF_3 modification of different tryptic peptides with annotation of calculated SAS for each reactive residue.

To evaluate the impact of trifluoromethylation on a protein bearing numerous Trp residues, the structure and activity of Ly was evaluated. CD spectroscopy revealed no changes to the secondary structure of Ly modified under standard conditions - Fig. 2.20a. Careful quantification of the protein concentration after trifluoromethy-

lation and buffer exchange revealed that $\sim 80\%$ of the sample could be recovered - Fig. 2.20b.



	Activity [U]	**Specific activity** [U/mg]	**Relative activity** [%]
Ly	35.7 ± 3.7	181600 ± 18700	100 ± 9.7
Ly_CF_3	26.4 ± 2.5	178400 ± 17200	98.2 ± 9.4

Fig. 2.20: **a**| Comparison of CD spectra for native and modified under standard conditions **b**| Micro Bradford assay quantification of protein concentration before and after the trifluoromethylation reaction. The same samples were used in **c**| a *Micrococcus lysodeikticus* hydrolysis assay to determine the enzymatic activity. The results of the assay confirmed no change in the specific activity of Ly.

A standard hydrolytic activity assay, in which the rate of *Micrococcus lysodeikticus* cell

wall lysis (due to the structural peptidoglycan hydrolysis) was quantified, confirmed the retained enzymatic function of Ly_CF_3 - Fig. 2.20c. These results combined, showed that even on a protein with many Trp residues, the CF_3 modification has a minimal effect on the structure and function.

2.4 Proof-of-principle PrOF experiments

Having successfully developed a methodology for the CF_3-ylation of proteins, containing various numbers of Trp residues, and carried out an extensive characterisation, the next aspect of the project was to highlight the utility of the modified protein products. This was deemed important in order to highlight that in certain types of experiments it is possible to use a heterogeneously fluorinated product mixture, without the complete knowledge of which specific residues are modified. Although the peak assignment increases the amount and the resolution of information that can be obtained in a protein observed ^{19}F NMR experiment (PrOF), there are some applications, such as a ligand screening assay or general structural assessment, where it is not a necessary requirement.

2.4.1 Annexin V - ligand state

^{19}F NMR experiments aimed at establishing the selectivity of the A5 trifluoromethylation concluded that the reaction conditions resulted predominantly in modification of the plethora of Tyr residues present in the protein. However, it also revealed partial fluorination of the single Trp residue, likely in a homogeneous population, since the LC-MS analysis indicated that the majority of the protein bore a single CF_3.

Interestingly, in the native spectrum of the A5_CF_3 no distinct peak corresponding to the lower intensity Trp modification was observed.

The proteins from the annexin family characteristically bind to phospholipid membranes in a Ca^{2+} dependent manner, with an important contribution from the region containing the single Trp residue[256]. A comparison of the crystal structures for the proteins in two different calcium induced conformational states suggested a high conservation of the overall fold[257], with an important exception of the region surrounding Trp187 - Fig. 2.21a. Therefore, it was speculated that the addition of this cation would induce a conformational change in the A5 protein that should be reported by the fluorinated Trp residue (with the Tyr signals essentially unchanged) and manifested as a small change in the ^{19}F NMR spectrum. Indeed, the addition of 10 mM $CaCl_2$ induced the appearance of a new distinct peak that disappeared upon denaturation of the protein - Fig. 2.21b. The low signal intensity suggested that it was due to the Trp residue, however, it is unclear why it could not be seen in the native spectrum. Possibly, due to the unfavourable relaxation properties (broadness) or an overlap with the Tyr peak. This result suggested that even under conditions of incomplete Trp selectivity, the established trifluoromethylation protocol can be used for PrOF experiments.

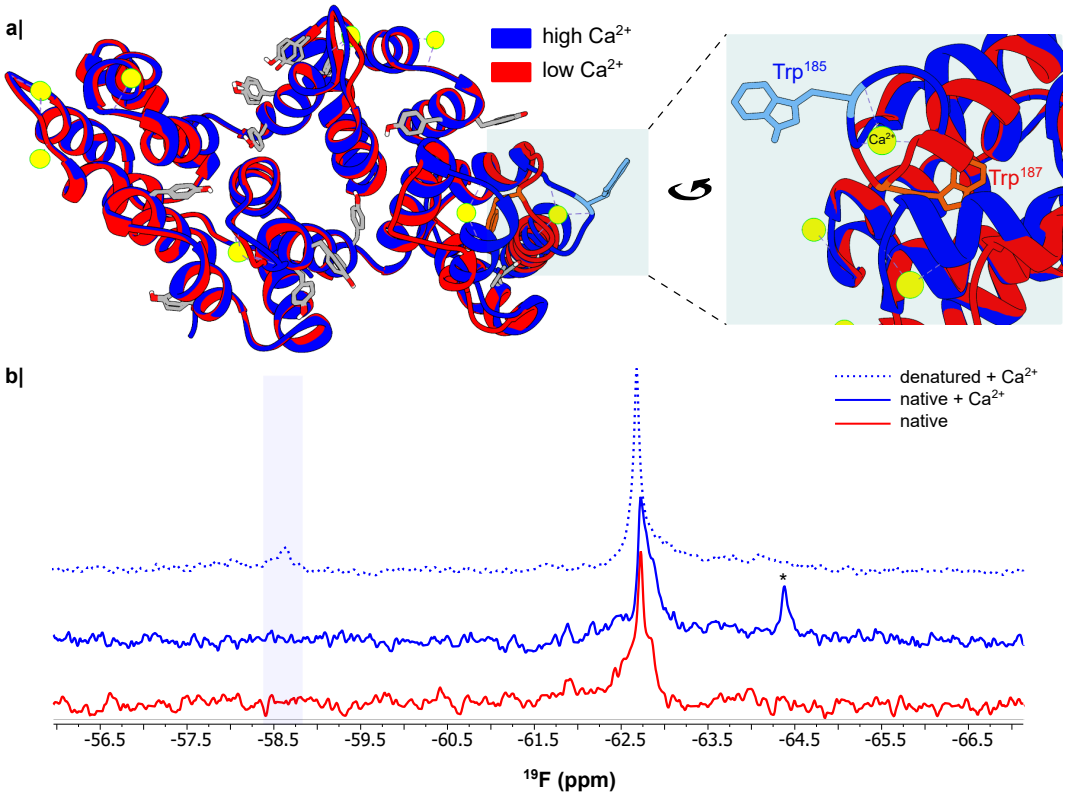


Fig. 2.21: **a**| Visualisation of rat A5 in the presence of high Ca^{2+} concentration (blue), overlaid with the crystal structure of human A5 obtained under low Ca^{2+} conditions (red) showing conformational change, and position of the Trp residue, PDB (2RAN and 1AVR)[257] **b**| Comparison of fragments of the ^{19}F NMR spectra of native A5_CF_3 (solid red), in the presence of 10 mM $CaCl_2$ (solid blue) with a new peak marked with an asterisk, and calcium sample denatured with 6 M urea (dotted blue) ($\sim$ 100 μM protein in 100 mM NH_4OAc pH 6.0), with region of expected **TrpCF$_3$** peak highlighted.

2.4.2 Myoglobin - ligand screening

A ligand screening experiment was carried out for eqMbCF$_3$ after verification that the trifluoromethylation reaction did not cause the displacement of the water molecule bound by the heme (Fig. 2.13). The protein sample was exposed to different ligands and ^{19}F NMR spectra were measured to evaluate the impact of the (potential) binding event. First, established ligands (cyanide - CN^- and imidazole - Im)[258] produced a

significant change - Fig. 2.22a, with a stronger effect on the more upfield peak assigned to Trp14, presumably due to its tighter packing as discussed in Fig. 2.10. Interestingly, another known ligand - fluoride - F^-, did not cause signal perturbation, suggesting a minuscule impact on the protein fold, possibly due to its relatively small size. Sarcosine (Sar, N-methylglycine) and trehalose (Tre) that serve as an osmoprotectant and/or a cryoprotectant[259] agents promoted a global change with both signals being shifted to a similar extent. Such behaviour corresponded to their non-specific mode of interaction, affecting the total character of the solvent.

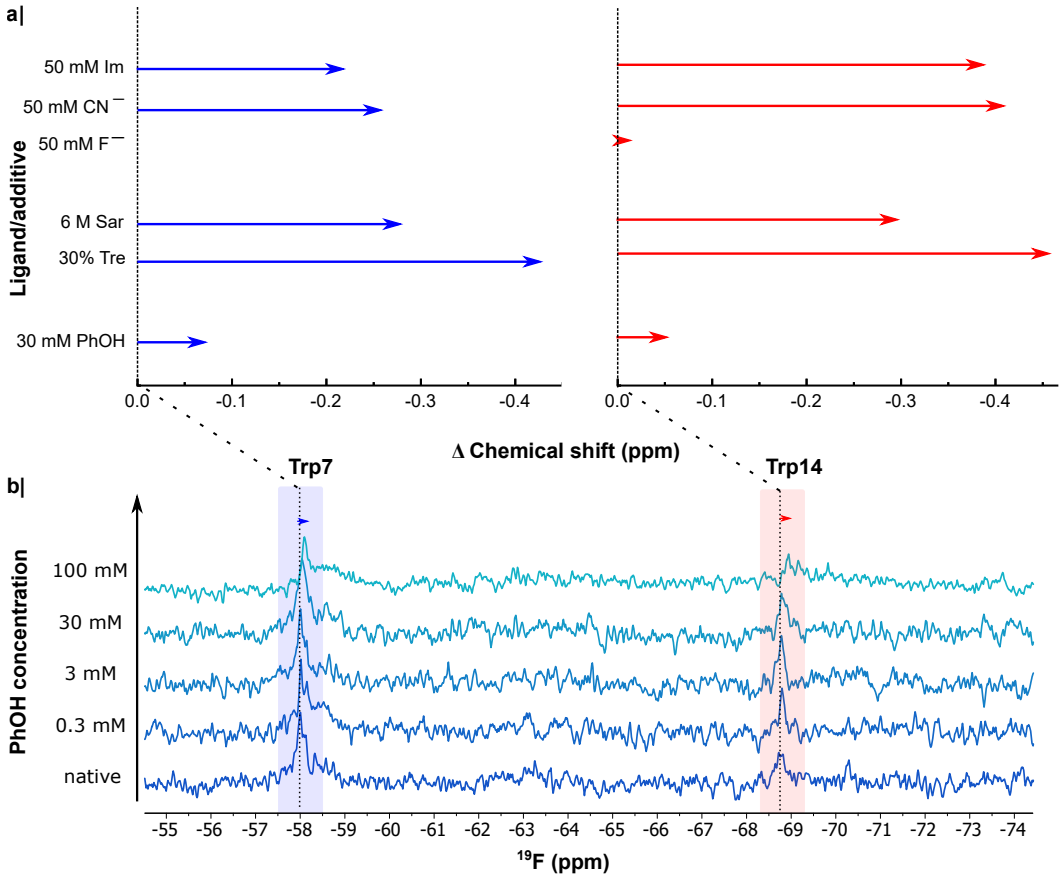


Fig. 2.22: **a**| The chemical shift change of eqMbCF$_3$ ($\sim (100-300)$ μM in 100 mM NH_4OAc pH 8.0) Trp peaks induced by the presence of various ligands and additives (after 4 h incubation at ambient conditions). **b**| Fragments of the ^{19}F NMR spectra of eqMbCF$_3$ titrated with phenol (PhOH) showing low response to this compound.

Finally, a putative substrate - phenol (PhOH) was tested[260]. However, in a range of concentrations, only a very small change was observed, without a clear concentration dependent response ruling out specific heme binding - Fig. 2.22b.

2.4.3 Lysozyme - ligand binding quantification

In the next experiment, a ligand binding quantification assay was carried out for lysozyme (6 Trp residues). For a proof-of-concept experiment a ligand with an average binding affinity (low $\sim\mu$M) was chosen - alkaloid berberine[261], a good representative of a pharmaceutical fragment hit. PrOF titration assay was carried out[223], which resulted in obtaining a series of differential ^{19}F NMR spectra - Fig. 2.23a. For the analysis, the most intense peak was chosen (-57.8 ppm), the center of which was determined using a generalised Lorentzian fitting function (MestReNova v11). In order to estimate the strength of the binding, first the differences in the chemical shift needed to be corrected for the DMSO induced shift by applying a linear correction function. This allowed fitting of a binding isotherm that yielded a value of $K_{\mathrm{d}} = (20 \pm 18)$ μM, which was in a good agreement with the known binding constant of this compound (16.8 ± 0.7) μM determined by isothermal titration calorimetry (ITC)[261] - Fig. 2.23b. Above proof-of-concept experiments demonstrated the potential of the developed trifluoromethylation technique to prepare fluorinated proteins applicable for various straightforward NMR experiments.

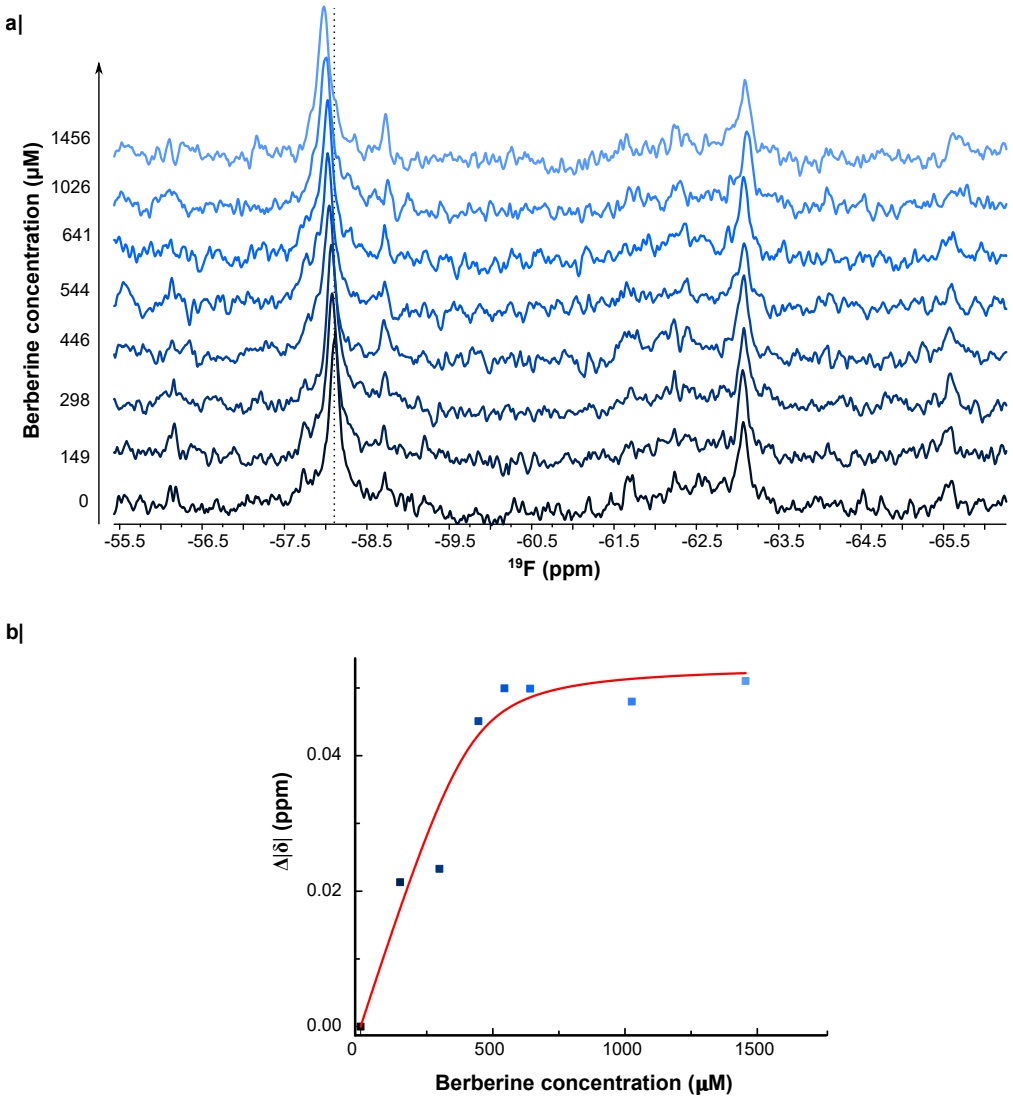
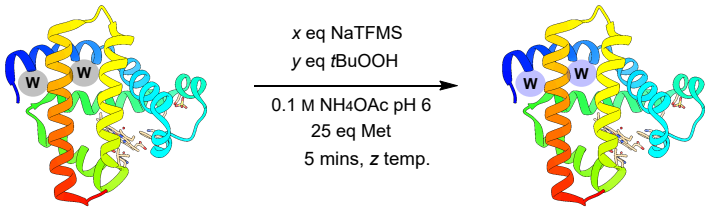


Fig. 2.23: **a**| Fragments of ^{19}F NMR spectra for Ly_CF_3 (420 μM in 100 mM NH_4OAc pH 6.0) during titration with berberine (50 mM stock in DMSO). **c**| Corrected chemical shifts differences and ligand concentrations were plotted and the dependence fitted with a binding isotherm (T = 25 °C) using a non-linear curve (Origin 2017), with a fixed protein concentration and the rest of the parameters free.

2.5 Perspectives for radiofluorination

Considering the short reaction time and operational simplicity of the ^{19}F trifluoromethylation protocol developed herein, a question was posed as to whether an

adaptation for ^{18}F trifluoromethylation would be possible. The main challenge in developing radiofluorination methodologies is the low concentration of the unstable isotope available. As such, the limiting reactant is always the labelling reagent - in this case the radical precursor. This posed an immediate problem, since, relative to the protein sample, the optimised protocol uses a large excess of the radical precursor. In order to test if the translation might be feasible, a range of conditions were screened, with an increased biomolecule amount and a progressively lower excess of NaTFMS. At a low protein concentration, the trifluoromethylation was not observed in the desired timescale (∼mins), requiring at least >5 mM of the labelling reagent (Fig. 2.24 entries 1-3). When the protein concentration was increased 20 fold (to the solubility limit), after 5 mins, the reaction yielded trifluoromethylated product with ∼20% conversion with respect to the radical precursor, under almost stoichiometric ratios (Fig. 2.24 entry 6). Even though the kinetics of the reaction is likely to be complex[178], for the studied protein trifluoromethylation one may estimate a pseudo-second-order rate constant as $\sim 10^{-1}$ M^{-1} s^{-1} , which falls within a range typically encountered for the successful radiofluorination procedures[262]. It was deemed likely that the protocol would need additional optimisation under radiochemistry conditions. However, taken together, it was concluded that the protocol may be adaptable for protein radiofluorination.



eqMb
57 or 1100 μM

Entry	NaTFMS [eq]	*t*BuOOH [eq]	Temp. [°C]	$+CF_3$
		eqMb at 57 μM		
1	200	25	rt	1.6
2	100	25	rt	0.4
3	50	25	rt	none
		eqMb at 1.1 mM (20x higher)		
4	10	2.5	40	>4
5	5	1.25	40	1.3
6	5	1.25	rt	1
7	1	2	rt	traces
8	1	2	40	traces

Fig. 2.24: Trifluoromethylation of eqMb with variable amounts of reagents and temperature with LC-MS quantification of the modification copy-number. Lowering the amount of radical precursor, as expected, required an increase in the protein concentration in order to observe appreciable conversions.

2.6 Discussion and conclusions

This chapter has described a novel protein trifluoromethylation protocol, which was first investigated on a small molecule and peptide model. These experiments were important to establish that this radical transformation shows preference for Trp residues. During the course of, and since, this work was published, several studies on the biomolecules trifluoromethylation utilising NaTFMS as a radical source, have appeared in the literature[263] - Fig. 2.25. In one of them, the authors acknowledged the

superior Trp reactivity, but instead capitalised on modification of the second most reactive residue - Tyr[194]. Optimisation of the reaction conditions (higher temperature, lower pH) and exploration of a different initiation mode (photoredox), resulted in a successful Tyr trifluoromethylation protocol, efficient in absence of Trp residues.

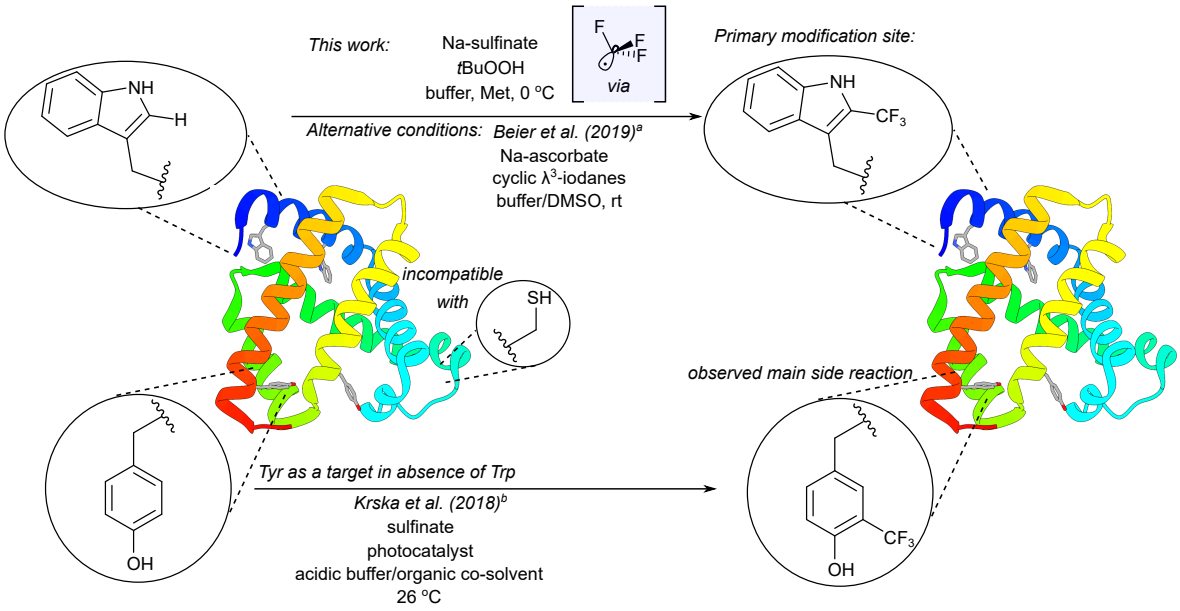


Fig. 2.25: Overview of the recent biomolecule trifluoromethylation methodologies and comparison with this work. [a] - alternative radical precursor[123], [b] - alternative conditions for sulfinate activation[194].

The photoredox and metal catalysed reactions using NaTFMS have also been optimised for the Trp trifluoromethylation in peptidic substrates, again confirming the selectivity conclusions which were observed in this study. However, useful conversions were only obtained using organic solvents[233,234]. In a broader study, focused on the fluoroalkylation of peptides with sulfinates, the same reactivity/selectivity patterns emerged for the trifluoromethyl radicals[78]. Radical trifluoromethylation of peptides and proteins was also accomplished using a different radical source (fluoroalkyl-substituted cyclic λ^3-iodanes). Pleasingly, the obtained conclusions were aligned with

this thesis results, highlighting the versatility of this type of chemistry[123]. The authors of the study highlighted the multiple additions of the CF_3 group to Trp as an important side reaction in the presence of a high excess of reagents, most likely due to a higher efficiency of the radical generation.

Another interesting recent fluoroalkylation is the explicit electrophilic Cys trifluoromethylation in peptides, also applicable in radiochemistry settings[264]. Such studies are important in the view that developed here radical trifluoromethylation was found to be incompatible with free Cys. Since then, the idea of the biocompatibility of fluoroalkyl radicals has spread and essentially identical methodology has also been applied to different biomolecules (nucleotides)[265]. Summing up, the developed protocol, taking advantage of unique features of radical chemistry, unequivocally allowed activation of the C2 position of Trp residues, which prior to this was only robust for peptidic substrates where the N-H influence is limited by use of a protecting group[266]. A range of converging approaches, with a variety of conditions and reagents, provides a further advantage for a protein chemist trying to achieve protein trifluoromethylation.

With the reactivity and selectivity data in hand, the conditions were optimised for protein trifluoromethylation. An in-depth study of model protein trifluoromethylation helped to elucidate the most important features of the reaction such as lack of requirement for the residue solvent exposure and therefore good access to buried sites. This mode of reactivity is especially attractive, as $\sim$90% of the Trp residues are found in the interior of proteins, with low solvent accessibility[267].

In this project, a very small set of experiments was dedicated to characterise the

detailed structure and function of proteins with trifluoromethylated native aromatic residues. This limited study did not reveal significant impact of such a modification. However, this may be somewhat surprising, as it is well known that even a single substitution of C-H bonds with C-F counterparts can have a dramatic effect on the amino acid polarity and secondary structure propensity, in turn leading to differential protease stability, folding and map of local interactions for the protein incorporating it[268]. For example, it has been shown that the introduction of fluorinated Phe analogues into amyloidogenic peptides significantly affects kinetics of the aggregation process[269,270]. However, because the developed here trifluoromethylation procedure is applied to natively folded proteins, it is unclear how the accommodation of the CF_3 group may influence protein stability and fold, with the outcome most likely to be protein dependent. Little is known about the general impact of the non-analogous amino acids with the trifluoromethyl group positioned away from C(α) (as in TrpCF_3) and this area requires a more systematic study to elucidate it.

As the introduction of trifluoromethyl group not only affects the hydrophobicity but also the electron distribution, it may lead to the formation of new non-covalent interactions and may exert profound effect on protein-protein interactions involving modified residues, as it was demonstrated in the literature[205]. Since Trp residues are often found on interfaces of such interactions[271], it would be interesting to see how trifluoromethylation may be used to modulate them.

The key part of the investigation was the extension of substrate scope. The protocol turned out to be generally applicable to proteins with various sizes, folds and numbers of Trp residues. However, not unexpectedly, the detailed selectivity of

the reaction turned out to be highly protein dependent, an important consideration when deciding on the fluorination method. The selectivity evaluation is a simple experiment, useful as well for the further PrOF studies. Fully Trp selective reactions were possible for certain proteins, with others displaying partial modification of His and Tyr residues. However, by performing proof-of-concept NMR experiments it was shown that also in these cases the constructs could still be useful for various ^{19}F NMR experiments. Considering the general abundance of the amino acid residues in proteins, the average Tyr to Trp ratio is 2.2 (in the human proteome)[272]. Such a value, based on the examples in this study, should provide an optimal Trp selectivity for an 'average' protein and be conducive to widespread adoption of this PrOF enabling protein fluorination methodology.

2.7 Experimental

2.7.1 General information

Reagents and solvents were purchased from commercial suppliers (Sigma-Aldrich UK, Acros Organics UK, Alfa Aesar UK and Fisher Scientific UK) and used without further purification. Proteins used in this study were sourced as followed:

Tab. 2.4: Identification of proteins used in the study and their source.

Abb.	**Name**	**Species**	**UNIPROT**	**Source**
A5	annexin V	*Pan troglodytes*	ANXA5_PANTR	In house expression
PanC	pantothenate synthetase	*Mycobacterium tuberculosis*	PANC_MYCTU[a]	In house expression[b]
SRP19	signal recognition particle	*Homo sapiens*	SRP19_HUMAN[c]	In house expression
eqMb	heart myoglobin	*Equus caballus*	MYG_HORSE	Sigma-Aldrich (M1882)
Tr	pancreas trypsin	*Sus scrofa*	TRYP_PIG	Sigma-Aldrich (93615)
Ly	egg white lysozyme	*Gallus gallus*	LYSC_CHICK	Sigma-Aldrich (62971)

[a] - with C-terminal His tag (KLAAALEHHHHHH), [b] - performed by Dr Jitka Dadova
[c] - with N-terminal His tag (MGSSHHHHHHSSGLVPRGSH)

Wild type PanC was sourced from Dr Jitka Dadova (Ben Davis group), the expression and purification procedure for this protein has been previously described[273]. All protein reactions were carried out in 2 mL microcentrifuge tubes open to air, without stirring. For all experiments, ultrapure water prepared with a Direct-Q^{TM} 3UV system (Merck - resistivity >18.2 MΩ· cm), additionally passed through a 0.22 μm Millipore filter, was used. Room temperature (rt) refers to ambient temperature,

while PBS refers to Dulbecco 'A' phosphate buffered saline i.e. 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 137 mM NaCl, 3 mM KCl, pH 7.3 prepared according to the manufacturer's instructions with Oxoid^{TM} Phosphate Buffered Saline Tablets (Thermo Scientific). For the small scale buffer exchanges PD SpinTrap^{TM} (up to 140 μL), PD MiniTrap^{TM} (0.5 mL) and PD MidiTrap^{TM} (1 mL) desalting columns with Sephadex G-25 resin (GE Healthcare Life Sciences) were used unless stated otherwise. Thin layer chromatography (TLC) was performed using Merck Milipore aluminum TLC plates, silica gel coated with the fluorescent indicator F254. The plates were visualised with a UV lamp and by reaction with ninhidrin dip (1.5 g ninhydrin in 100 mL of *n*-butanol and 3 mL acetic acid) with a subsequent heat gun development. For flash column chromatography, Geduran^{TM} Si 60 (0.040-0.063 mm) silica gel was used. HPLC purification of small molecules was performed using Shimadzu 20 Series HPLC equipped with the Phenomenex Synergi^{TM} 4 μm Hydro-RP (80 Å 100 x 21.20 mm) reversed phase C18 polar capped column. Elution was performed with an optimised gradient program (0 min $\rightarrow$ 5% B', (0-2) min $\rightarrow$ 5% B', (2-20) min $\rightarrow$ 40% B') using as A': 0.1% formic acid (FA) in H_2O and as B' (0.1% FA in MeCN) and total flow of 12 mL/min. The stock solution of porcine trypsin was prepared with 10 mM calcium chloride and kept on ice to reduce autoproteolysis. Most of the experiments were carried out as single experiments ($n = 1$) under given conditions. Therefore, all reported yields and conversions have to be treated semi-quantitatively. This has been conducted bearing in mind that numerous comparable reactions/experiments were performed all supporting general trends. In the case of protein LC-MS/MS experiments, where higher experimental uncertainty was observed, the experiments were

performed at least in duplicate.

Protein characterisation

Protein concentration was determined either by a measurement of a UV-VIS absorbance using NanoDrop™ ND-1000 (Thermo Fisher Scientific) (myoglobin ε_{409} = 157 000[274], PanC ε_{280} = 15 930[273] and for other proteins approximate values calculated with the Expasy ProtParam tool[275] for a given amino acid sequence from the UNIPROT database: LYSC_CHICK, ANXA5_PANTR, TRYP_PIG, verified with LC-MS) or by performing the Bradford assay with the 96-well plate reader SpectroStar™ Omega (BMG LabTech) according to the manufacturer's instructions (on a 15/185 v/v protein sample to Bradford reagent ratio).

Before loading onto the gel for SDS-PAGE, protein samples were denatured with 5X Laemmli buffer (CSH Protocols) and incubated at 95 °C for 15 min. The analyses were done using precast NuPAGE™ 4-12% Bis-Tris Protein Gels (Invitrogen) in MES running buffer prepared from 20X Bolt™ stock solution (Invitrogen). SeeBlue™ Plus2 Pre-stained Protein Standard (Invitrogen) was used as a mass marker. Electrophoresis was carried out at 200 V for 40 mins in MiniCell™ tanks (Novex) and a PowerPac™ controller (Bio-Rad). After that the gels were visualised with the InstantBlue™ stain (Expedeon Ltd.) for ∼60 minutes. Gels were destained by 3x washing with water (15 min, 1 h and overnight) prior to digital scanning.

Protein expression

Culture media and glassware used in the preparation of bacterial cultures were sterilised prior to use by autoclaving for 20 minutes at 121 °C in a Touchclave-R™

autoclave (LTE Scientific). Handling of the cultures was performed in an A-range laminar flow workstation (BASSAIRE). Growth of bacterial cultures was realised in an incubated shaker (Innova 42R). Centrifugation of bacterial cultures was carried out in an Avanti^{TM}-JXN-30 centrifuge (Beckmann Coulter) with a JLA9-1000 rotor (Beckmann Coulter, 1 L bottles), while cell lysates were spinned down in the Heraeus^{TM} Megafuge^{TM} 40R benchtop centrifuge with a Fiberlite^{TM} F15-6 x 100y fixed-angle rotor (Thermo Scientific) using volumes up to 50 ml in the centrifuge tubes (Greiner Bio-One). Small volume variable temperature incubations were performed in a benchtop PHMT-PSC18 thermo-shaker (Grant Bio) while centrifugations in a benchtop 5427R centrifuge (Eppendorf). Sonication was performed with a Model 505 Sonic Dismembrator (Fisher Scientific). Proteins were purified with an Äkta^{TM} FPLC System UPC-900 (GE Healthcare). Plasmid sequencing was confirmed by sequencing services of Oxford University using the DNA Sanger Sequencing technology (Source Bioscience).

NMR Spectroscopy

Deuterated water (small molecules) or NH_4OAc (100 mM pH 6.0 or 8.0) prepared in D_2O were used for NMR spectroscopy. Chemical shifts are reported in parts per million (ppm) relative to the centre of the reference signal: for ^1H: D_2O (4.79 ppm) and for ^{13}C: added CH_3COOH (21.03 ppm). Fluorine spectra were referenced against the residual sodium triflinate signal (−88.45 ppm), the chemical shift of which was determined relative to that of trifluoroacetic acid (−76.5 ppm versus $CFCl_3$). Spectra were analysed using the Mestrelab software (MestReNova v11.0.2). Relevant 2D experi-

ments were performed to facilitate the signal assignment. All NMR experiments for small molecules were performed on a Bruker AVIIIHD 400 nanobays (400.1 MHz for proton) running TOPSPIN 3 equipped with a 5 mm z-gradient multinuclear BBFO probe. Protein ^{19}F NMR spectra were typically acquired on a 11.7 T (500 MHz) spectrometer, at 298 K with acquisition times of (0.4 – 1) s and recovery delays of (0.4 - 1) s. Multiplicity is denoted by the bracketed symbol. Coupling constants (J) are reported to the nearest 0.1 Hz. Quantitative fluorine experiments were performed using NMR grade 2,2,2-trifluoroethanol (−77.7 ppm, t, 3F) as the internal standard with following experimental parameters: number of scans: 128 - 256, acquisition time: 3 s, delay time 7 s and with resolution not smaller than 0.33 Hz/pt. Thermal denaturation and titration experiments were performed in collaboration with Dr Gogulan Karunanithy and Prof. Andrew Baldwin (PTCL, University of Oxford). Registration of each spectrum was followed by a 10 mins equilibration period.

Mass Spectrometry

Small molecule high-resolution mass spectra were acquired on a Bruker Micro TOF-ESI mass spectrometer. LC-MS analyses for proteins used Waters LCT Classic MS coupled with Shimadzu 20 Series HPLC. For desalting, Merck Chromolith^{TM} C18 (50 x 2 mm) or Thermo Proswift^{TM} (250 x 4.6 mm x 5 μm) columns were used. H_2O:MeCN (95:5, solvent A) and MeCN:H_2O (95:5, solvent B), both with 0.1% formic acid, were the mobile phases with a flow rate of 0.4 mL/min. The standard LC gradient program was: 0 min → 100% A, (0-2) min → 100% A, (2-8) min → 100% B, (8-10) min → 100% A. The following MS instrument parameters were used: capillary

voltage 3 kV, sample cone voltage 100 V, desolvation temperature 300 °C, source temperature 120 °C, cone gas flow 50 L/h, desolvation gas flow 300 L/h (both N_2). The ion series was calibrated to at least 19 matched peaks of the multiply charged ion series of eqMb analysed in the same way.

Tandem Mass Spectrometry

Tryptic peptides were analysed on an NanoAcquity^{TM}-UPLC system (Waters) coupled to an Orbitrap^{TM} Elite MS possessing an EASY-Spray^{TM} nano-electrospray ion source (Thermo Fischer Scientific). The peptides were trapped on a packed guard column (75 μm i.d. x 20 mm, Acclaim PepMap100 C18, 3 μm, 120 Å) using solvent A'. Peptides from the injection (2 μl) were separated on an EASY-spray Acclaim PepMap^{TM} analytical column (75 μm i.d. x 15mm, RSLC C18, 3 μm, 100 Å) using a linear gradient (length: 68 mins, 3% $\rightarrow$ 97% solvent B', flow rate: 300 nL/min). Full-scan mass spectra were collected (scan range from 350 to 1500 m/z, resolution 120 000, automatic gain control (AGC) target 1E6, maximum injection time 100 ms). After each MS scan, the 10 most intense peaks were selected for CID fragmentation at 35% of normalised collision energy. CID spectra were acquired in the Iontrap^{TM} (AGC target 1e3, maximum injection time 50 ms).

MS/MS Data Analysis

Raw data files were processed using the ThermoBioPharma software (Version 2.0) with a Default Peptide Mapping Method. FASTA sequences of used proteins were downloaded from UniProt database (March 2017) or supplied manually. A maximum number of 3 missed cleavages for digestion enzymes was chosen. Oxidation (MW) and

trifluoromethylation (for WYCHFM $\Delta m = +68$) were used as variable modifications. In case of Cys containing protein carbamidomethylation ($\Delta m = +57$) was chosen as a fixed modification. All assigned spectra were manually validated. The relative modification was reported here is the '*%Abundance*' parameter calculated by the software.

Analysis of LC-MS data for proteins

In a typical LC-MS analysis for monitoring of a protein reaction an aliquot of reaction mixture was diluted with H_2O to 0.05 mg/mL protein concentration. Such a solution was injected (5 μL) for the analysis during which the total ion count (TIC) was monitored producing a chromatogram, with an example shown in Fig. 2.26a. The mass spectra from the discernible protein peak ($\sim$ 7.4 min) were combined to produce an unprocessed ion series - Fig. 2.26b. The spectra were cut to include only the visible envelope of the protein ion series and the total mass was reconstructed using the MassLynx Max(imum) Ent(ropy) deconvolution algorithm (Waters) - Fig. 2.26c. Protein molecular weights were calculated with the ExPasy ProtParam tool (http://web.expasy.org/protparam/). The following deconvolution parameters were used: Output mass range: $M \pm 8000$ Da with 1 Da/channel resolution, where M is the expected molecular mass of the protein. The damage model assumed the uniform Gaussian model with the width at half weight determined according to that of the most intensive protein peak.

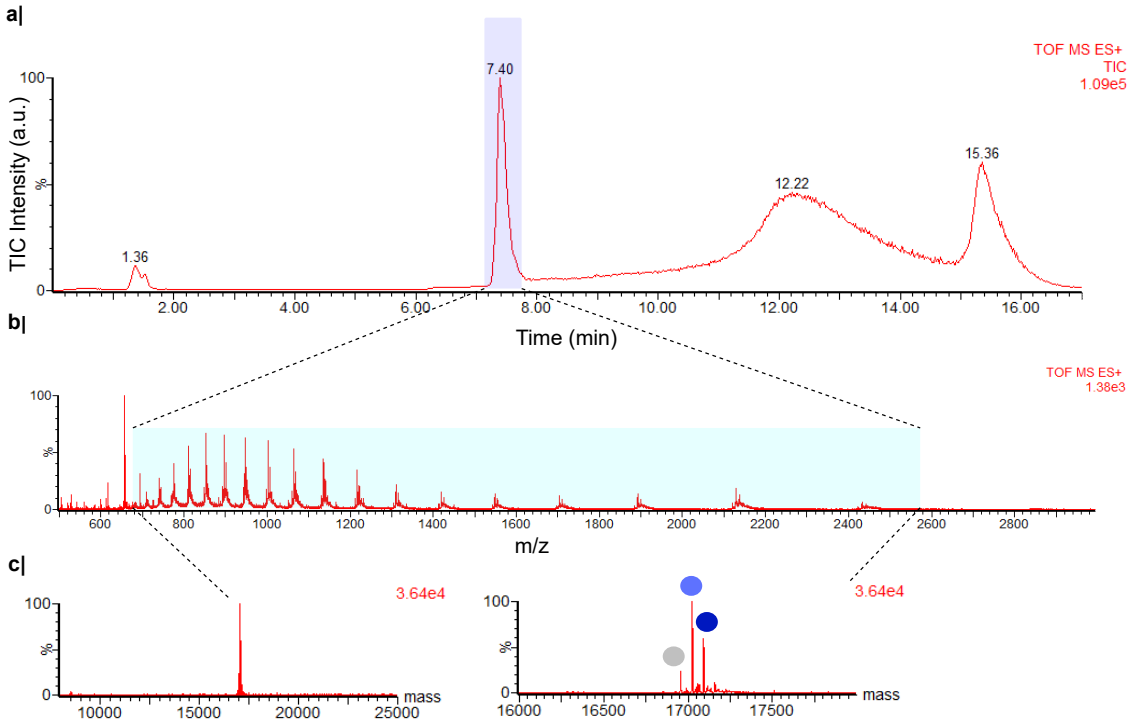


Fig. 2.26: **a**| Total ion count chromatogram of a typical eqMb trifluoromethylation sample analysis with the protein peak highlighted in dark blue. **b**| Combined mass spectra from the protein peak, the m/z range for deconvolution is highlighted in light blue. **c**| Reconstructed protein mass spectra with Max Ent 1 algorithm, shown different mass ranges.

33% was used as the minimum intensity ratio and the algorithm was iterated to convergence. Conversions to the modified products were determined by calculation of intensity ratios between a peak of interest and intensity sum for all of the peaks. Such a method has been previously applied for the quantification of protein reactions mixtures. It was concluded that ionisation of proteins (and therefore observed peak intensity) does not change because of addition of relatively small modifications, as the process is dependent on the total surface composition[47,53,276].

Spectroscopy

CD spectroscopy was carried out using a Chirascan^{TM} CD spectrometer (Applied Photophysics) in a Peltier temperature-controlled cell holder. Experiments were performed in duplicate at 25 °C in a 0.1 cm path length cuvette using $\sim$ 0.2 mg/mL protein in 10 mM PBS (pH 8.0) unless stated otherwise. Spectra were recorded from 185 to 260 nm with 0.5 nm intervals. The data for each point was averaged for 2 s. Obtained spectra were baseline corrected and normalised at the point of maximum elipticity for the comparison. UV-VIS spectra were acquired on SpectroStar Omega spectrometer (BMG LabTech). Infrared spectra (IR) were acquired with Bruker Tensor 27 equipped with a diamond attenuated total reflection (ATR) system.

2.7.2 Experimental procedures

Natural AAs reactivity and selectivity screening

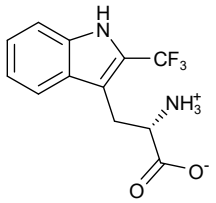
An amino acid (31 μmol, (2.3 - 6.3) mg) was weighed into a microcentrifuge tube and dissolved in NH_4OAc buffer (457 μL, 1 M, pH 8.0 prepared in D_2O, uncorrected for deuterium). The mixture was mildly heated (40 °C) and sonicated for 5 mins to allow for the complete dissolution (Tyr was only partially soluble under these conditions as confirmed in a separate experiment) and then put on ice. Next, NaTFMS was added (sodium triflinate, 80 μL, 3.9 M prepared in the deuterated buffer). The reaction was initiated by the addition of tBuOOH (10.8 μL, 7.2 M in H_2O). After 2 hours, the reaction mixture was removed from ice and incubated at ambient conditions for a further 46 hours. After that, 2,2,2-trifluoroethanol (2.2 μL, 30 μmol, 95% NMR Grade) was added. The microcentrifuge tube was briefly vortexed and the solution

transferred into the NMR tube. The reaction mixture was analysed by ^{1}H NMR and ^{19}F NMR. The same procedure was followed for the competition assay.

Standard conditions for trifluoromethylation: amino acids

An amino acid (0.03 mmol, 1 eq) was weighed into a microcentrifuge tube and dissolved by mild heating (40 °C) and sonication (∼5 mins) with a minimum amount of NH_4OAc buffer (< 2 mL, 1 M, pH 6.0). The resulting solution was cooled on ice and NaTFMS (46.8 mg, 0.3 mmol, 10 eq) was added. The reaction was initiated by an addition of cooled *t*BuOOH (10.4 μL, 7.2 M aq. solution, 0.075 mmol, 2.5 eq). The mixture was kept on ice for 2 hours and then further incubated at rt for 46 hours. Next, it was frozen with liquid nitrogen and lyophilised. The crude solid was dissolved in a suitable solvent and subjected to purification.

Synthesis of TrpCF$_3$



Synthetic procedure according to the general protocol (0.09 mmol). The crude mixture was dissolved in DCM/MeOH/AcOH/H_2O (80:15:4:1) and purified via flash column chromatography using a linear elution gradient of MeOH in DCM/MeOH/AcOH/H_2O up to 65:20:10:5. **TrpCF$_3$** was isolated a white solid - acetate salt, ((40 ± 5)%, n = 3, ∼12 mg).

Rf: 0.74 (in DCM/MeOH/AcOH/H_2O - 65:20:10:5),

^{1}**H NMR** (400 MHz, D_2O) δ 7.77 (d, J = 8.1 Hz, 1H, H4), 7.53 (d, J = 8.3 Hz, 1H, H7), 7.38 (t, J = 7.6 Hz, 1H, H6), 7.24 (t, J = 7.6 Hz, 1H, H5), 4.02 (dd, J = 8.9, 5.9 Hz, 1H, Hα), 3.59 (dd, J = 15.3, 5.9 Hz, 1H, Hβ), 3.46 (dd, J = 15.3, 8.9 Hz,

1H, H'β),

^{19}**F NMR** (376 MHz, D_2O) δ −58.66,

^{13}**C NMR** (126 MHz, D_2O) δ 172.5 (COOH), 134.6 (C7a), 125.4 (C3a), 123.9 (C6), 122.5 (q, $^2J_{C,F}$ = 36.7 Hz, C2), 120.8 (q, $^3J_{C,F}$ = 267 Hz, CF_3), 119.7 (C5), 118.7 (C4), 111.5 (C7), 108.9 (q, $^1J_{C,F}$ = 2.7 Hz, C3), 54.3 (Cα), 24.4 (Cβ),

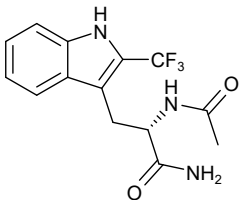
MS-ESI(+): $[M-H+2Na]^+$ found 317.0, calcd 317.0,

HRMS-ESI(+): $[M+H]^+$: calcd: 273.0845 found: 273.0845,

IR (ATR, $\tilde{\nu}_{max}$, cm^{-1}): 3185, 1619, 1507, 1393, 1321, 1253, 1163, 1114, 747,

$[\alpha]_D^{20}$ −32.9 (c = 0.2 in D_2O).

Synthesis of NATA-CF_3



Synthesis according to the general protocol but using as a solvent a 2:1 mixture of 1 M NH_4OAc pH 6 and DMSO (0.09 mmol). Crude mixture directly purified by RP-HPLC without freeze-drying. **NATA-CF_3** (N-Acetyl-2-trifluoromethyl-L-tryptophanamide) was isolated as a white solid ((12±4)%, n = 2, ∼3 mg).

Rf: 0.84 (DCM/MeOH 80:20),

^{1}**H NMR** (400 MHz, D_2O) δ 7.62 (dt, J = 8.2, 0.9 Hz, 1H, H4), 7.40 (dt, J = 8.3, 0.9 Hz, 1H, H7), 7.25 (ddd, J = 8.3, 7.1, 0.9 Hz, 1H, H6), 7.12 (ddd, J = 8.3, 7.1, 0.9 Hz, 1H, H5), 4.50 (dd, J = 7.8, 7.1 Hz, 1H, Hα), 3.32 (ddd, J = 14.2, 7.1, 1.1 Hz, 1H, Hβ), 3.16 (ddq, J = 14.2, 7.8, 1.1 Hz, 1H, H'β), 1.73 (s, 3H, CH_3),

^{19}**F NMR** (376 MHz, D_2O) δ −58.80 (s),

^{13}**C NMR** (126 MHz, D_2O) δ 175.7 ($CONH_2$), 173.9 ($COCH_3$), 135.6 (C7a), 126.9

(C3a), 124.9 (C6), 123.1 (q, $^{2}J_{C,F}$ = 36.8 Hz, C2), 122.1 (q, $^{3}J_{C,F}$ = 268.7 Hz, CF_3), 120.6 (C5), 119.9 (C4), 112.5 (C7), 111.6 (q, $^{1}J_{C,F}$ = 2.8 Hz, C3), 54.2 (Cα), 26.1 (Cβ), 21.8 ($COCH_3$),

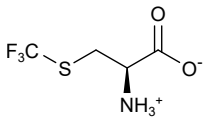
HRMS-ESI(−): $[M\text{-}H]^-$: calcd: 312.0965 found: 312.0965,

IR (ATR, $\tilde{\nu}_{max}$, cm^{-1}): 3232, 1637, 1417, 1317, 1230, 1164, 1112, 747,

$[\alpha]_D^{20}$ +42.4 (c = 0.17 in D_2O).

Synthesis of Cys-CF_3

Synthetic procedure according to the general protocol (0.03 mmol). After resuspension in water, the crude mixture was filtered to remove cystine and purified by ion exchange chromatography (DOWEX 50WX) eluting with 0.02% NH_3. After lyophilisation **Cys-CF_3** was isolated as a white solid - ammonium salt (13%, 0.8 mg).



^{1}**H NMR** (400 MHz, D_2O) 4.53 (dd, J = 8.2, 4.1 Hz, 1H, Hα), 3.64 (dd, J = 15.1, 4.1 Hz, 1H, Hβ)), 3.45 (dd, J = 15.1, 8.2 Hz, 1H, H'β),

^{19}**F NMR** (376 MHz, D_2O) δ −42.19 (s),

HRMS-ESI(−): calcd: 187.9999 found: 187.9998.

NMR spectra conforms to the literature data[277].

Expression and purification of Annexin V

The protein expression and purification protocol was adapted from literature[53]. A pET12a PAPI plasmid encoding the gene for annexin V was kindly given by Addgene company (originally deposited by Dr Jonathan F. Tait) and was used in the protein expression.

AA sequence (corresponding to wild type protein from *Pan troglodytes*):

AQVLRGTVTDFPGFDERADAETLRKAMKGLGTDEESILTLLTSRSNAQR
QEISAAFKTLFGRDLLDDLKSELTGKFEKLIVALMKPSRLYDAYELKHAL
KGAGTNEKVLTEIIASRTPEELRAIKQVYEEEYGSSLEDDVVGDTSGYYQ
RMLVVLLQANRDPDAGIDEAQVEQDAQALFQAGELK**W**GTDEEKFITIFG
TRSVSHLRKVFDKYMTISGFQIEETIDRETSGNLEQLLLAVVKSIRSIPAY
LAETLYYAMKGAGTDDHTLIRVMVSRSEIDLFNIRKEFRKNFATSLYSMI
KGDTSGDYKKALLLLCGEDD

MW_{calc} = 35805.6

A5 gene encoded in the plasmid was transformed into Novagen *E.coli*/BL21 (DE3) Single Competent cells. A single colony of bacterial cells was picked up and grown in LB medium at 37 °C overnight (8 mL with 100 mg/L ampicillin), then it was transferred into final culture in LB medium (2 mL per 1 L of LB with ampicillin) and grown at 37 °C until optical density (OD) at 600 nm of ∼0.6-0.7 was observed. Then, the protein expression was induced by an addition of IPTG (to final concentration of 1 mM) and cell cultures were incubated at 37 °C for an additional 4 h. Cells were harvested by centrifugation (7,000 rpm, 10 mins, 4 °C), the supernatant discarded and the pellet resuspended in the wash buffer (60 mL, 50 mM TRIS, 150 mM NaCl, pH

8.0). These suspensions were pelleted (15,000 rcf, 15 mins, 4 °C) and the supernatant discarded. Cell pellets were resuspended in $CaCl_2$ buffer (60 mL, 50 mM Tris , 10 mM $CaCl_2$, pH 7.2, 1 tablet of COmplete EDTA-free protease inhibitor cocktail/100 mL) and flash-frozen in liquid nitrogen for storage at −80 °C until the purification.

The cell pellet (from 2 L of the cell culture) was thawed on ice and DNase was added (2.5 mg), following an incubation on ice for 10 mins. The cell suspension was sonicated on ice (5 cycles, 50% amplitude, 30/60 s on/off) and the cell debris was collected by centrifugation (15,000 rcf, 15 mins, 4 °C). The supernatant was discarded, and the cell pellet resuspended in EDTA buffer (10 mL, 50 mM TRIS, 20 mM EDTA, pH 7.2). The cell suspension was centrifuged (15,000 rcf, 15 mins, 4 °C) and the supernatant collected. The cell pellet was washed with buffer W (20 mM TRIS, pH 7.8) and centrifuged (15,000 rcf, 15 mins, 4 °C). The pellet was discarded and the supernatant joined with the previous fraction and filtered with a 0.2 μm cellulose acetate membrane filter. The solution containing A5 was diluted with buffer W (up to 40 mL) and concentrated 8x using a centrifugal concentrator VivaSpin 20 (10,000 MWCO, 5,000 rcf, 4 °C). The procedure was repeated and the sample diluted with buffer W (up to 40 mL). The final solution (containing <0.5 mM of EDTA) was purified using anion exchange column chromatography (2 x 5 mL HiTrap Q, GE healthcare) with buffer W and a linear gradient of buffer E (20 mM TRIS, 500 mM NaCl, pH 7.8) - Fig. 2.27a, to afford the purified protein which was kept at 4 °C for short term use or frozen at −80 °C for long term storage. The protein was buffer exchanged before use and concentrated if needed (VivaSpin centrifugal concentrator MWCO 10,000). Yield from pure fractions: 20 mg/L

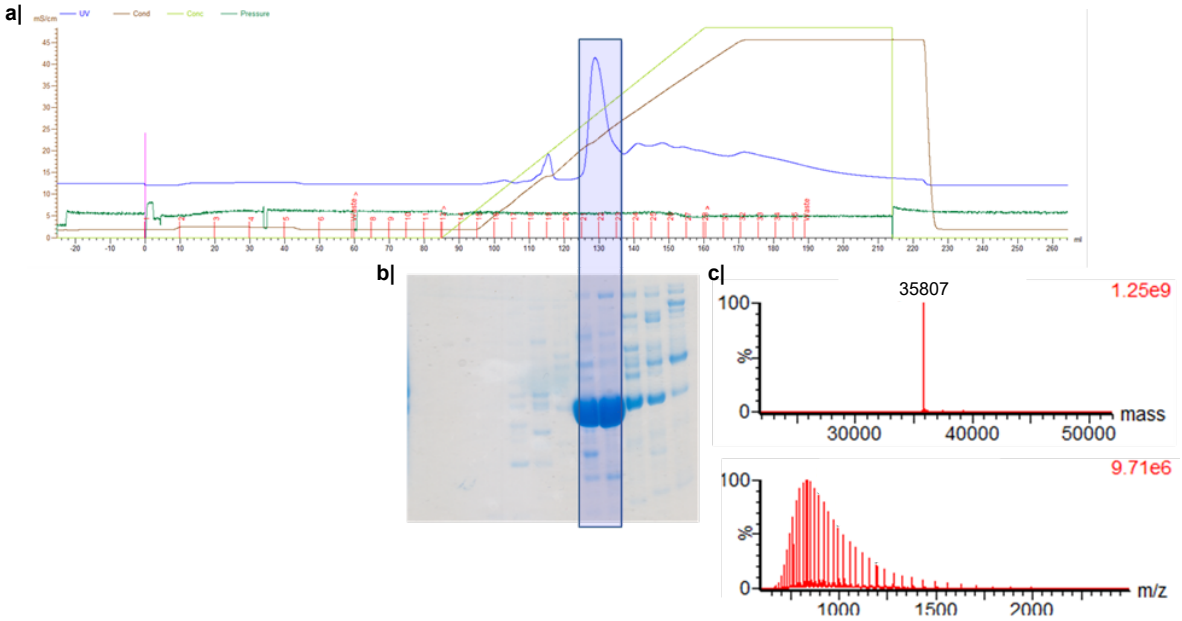


Fig. 2.27: **a**| UV chromatogram (blue trace) of an ion-exchange purification procedure, with a region of A5 pure fractions that were combined highlighted in blue. **b**| SDS-PAGE analysis of the collected fractions (corresponding with the UV trace above). Densitometry analysis for the combined fractions showed >90% purity. **c**| Reconstructed mass spectra and ion series of A5. The identity and purity of the protein was further confirmed by this analysis, which revealed a single species of the correct mass (±1).

Expression and purification of SRP19

A pET-15b plasmid encoding the gene for *Homo sapiens* signal recognition particle 19 (SRP19), transcript variant 1 was purchased from Genscript company (GenEZ ORF Clone: SRP19_OHu26556C_pET-15b) and used for the protein production. Protein expression and purification protocol was adapted from the literature[278].

AA sequence (corresponding to a His tagged human protein):

MGSSHHHHHHSSGLVPRGSHMACAAARSPADQDRFICIYPAYLNNKKTIA
EGRRIPISKAVENPTATEIQDVCSAVGLNVFLEKNKMYSRE**W**NRDVQYRG
RVRVQLKQEDGSLCLVQFPSRKSVMLYAAEMIPKLKTRTQKTGGADQSLQ
QGEGSKKGKGKKKK

MW_{calc} = 18188

BL21 (DE3) *E. coli* cells were transformed with the plasmid and grown on an LB agar (ampicillin) plate overnight at 37 °C. A single colony was picked up and propagated in LB media (1 mL with 100 μg/mL ampicillin). After 8 h incubation at 37 °C, an aliquot of this culture was added to 1LB to form the starting cultures (0.5 mL per 10 mL). Following an overnight growth, these were used to inoculate expression cultures that were grown at 37 °C until OD600 at 600 nm reached ∼0.6. Then, IPTG (up to 1 mM) was added and protein expression was carried out at 20 °C overnight. The cells were harvested by centrifugation (8,000 rpm, 10 min, 4 °C). The pellets were resuspended in lysis buffer (50 mL, 25 mM HEPES, pH 7.5, 1 M NaCl, 0.2 mM TCEP, 25 mM imidazole, 1 tablet of Complete EDTA-free protease inhibitor cocktail (Roche), 1 mg/mL lysozyme) and lysed by sonication (5 cycles, 50% amplitude, 30/60 s on/off). Cell debris was pelleted (15,000 rcf, 30 min at 4 °C) and the supernatant was filtered through a 0.45 μm filter and applied to 5 mL HiTrap Ni-NTA column (GE Healthcare) equilibrated in the purification buffer (25 mM HEPES, pH 7.5, 1 M NaCl, 0.2 mM TCEP, 25 mM imidazole). Following a wash (200 mL), the SRP19 protein was eluted with a linear gradient of elution buffer (25 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM TCEP, 250 mM imidazole). Fractions were analysed by SDS-PAGE and those containing pure protein of interest were pooled together, concentrated with VivaSpin 6 (MWCO = 5,000 Da) and dialysed or buffer exchanged into the desired buffer. The samples were stored at 4 °C for short term use or freezed at −80 °C until needed. Yield from pure fractions: 5 mg/L

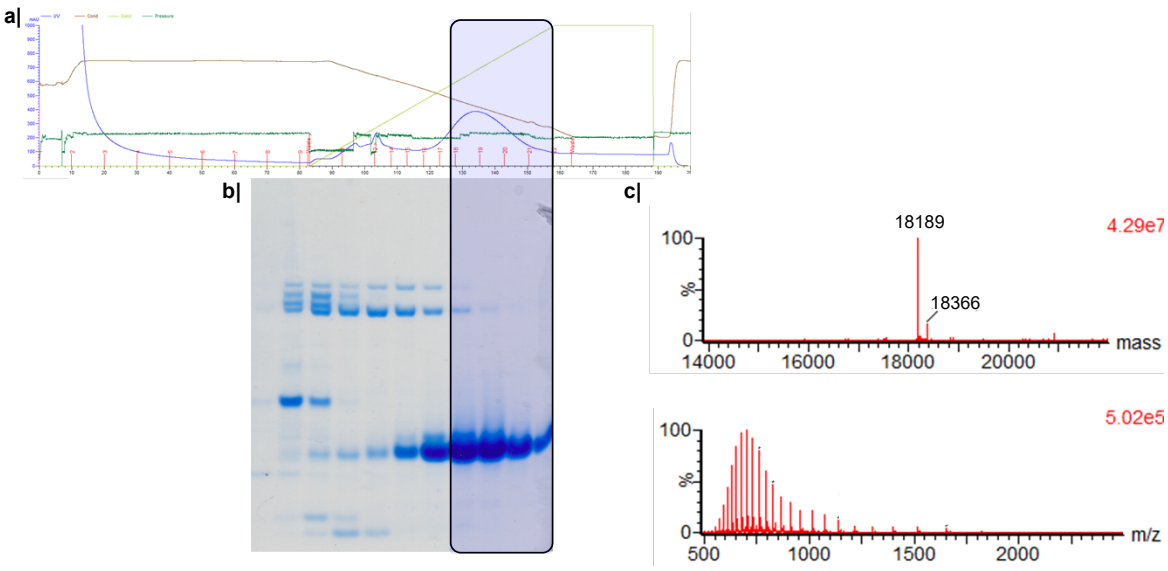


Fig. 2.28: **a**| UV chromatogram (blue trace) of a Ni-NTA affinity purification procedure, with a region of SRP19 pure fractions that were combined highlighted in blue. **b**| SDS-PAGE analysis of the collected fractions (corresponding with the UV trace above). A densitometry analysis for the combined fractions showed >95% purity. **c**| Reconstructed and ion series of mass spectra of SRP19. The identity and purity of the protein was further confirmed by this analysis, which revealed a peak of the desired mass (±1) alongside the less intense one (<10%) being an adduct corresponding most likely to gluconoylation due to the presence of sensitive N-terminal sequence MGSSHHHHHH[279] (Δm = +178, here +177).

Other protein handling and storage

Other commercially available proteins were stored at −20 °C as lyophilised solids, which were reconstituted in a desired buffer shortly before the experiment. PanC was stored at −80 °C as stock solutions in the storage buffer (1 mL, 6.7 mg/mL, 50 mM TRIS, 0.2% v/v βME, pH 6). Microcentrifuge tubes containing the protein were thawed in a beaker with water, transferred on ice and buffer exchanged into 100 mM NH_4OAc pH 6 for trifluoromethylation reactions.

Standard conditions for trifluoromethylation - low ∼μg scale

On an example of eqMb: NaTFMS (0.88 μL, 100 mg/mL, 200 eq), Met (0.53 μL, 20 mg/mL, 25 eq) and *t*BuOOH (7.1 μL, 10 mM, 25 eq) were added sequentially to eqMb (50 μL, 1 mg/mL = 57 μM, all solutions prepared in 100 mM NH_4OAc pH 6.0) in a microcentrifuge tube kept on ice. After the desired time, the reaction mixture was quenched with TEMPOL (30 μL, 0.4 mg/mL, ∼25 eq) and an aliquot (1.5 μL) was withdrawn for LC-MS analysis.

Standard conditions for trifluoromethylation - multi ∼mg scale

On an example of eqMb: A solution of eqMb (500 μL, 20 mg/mL in 1 M NH_4OAc buffer pH 6, 0.57 μmol) was cooled on ice, before a sequential addition of NaTFMS (17.8 mg, 200 eq), Met (2.1 mg, 25 eq) and *t*BuOOH (19.8 μL, 25 eq, 720 mM in the buffer). After the desired time, the reaction was quenched with TEMPOL (500 μL, 20 eq, 4 mg/mL in H_2O). The reaction mixture was desalted with a PD MidiTrap G-25 according to the manufacturer's instructions, diluted with D_2O (up to 10%) and directly used in ^{19}F NMR studies. For other proteins usually 2x - 3x lower scale (concentration) was used, with amounts of reagents scaled proportionally. If needed, before the ^{19}F NMR analysis the solution of protein was first concentrated with a VivaSpin (variable MWCO) centrifugal concentrators. Note different amount of reagents for proteins with a single Trp residue.

Average yield for: eqMb - (74 ± 8)% (Bradford assay), PanC - 64% (UV absorbance), Ly - (66 ± 12)%, A5 - (58 ± 8)% (UV absorbance), SRP19 - 50% (Bradford assay), Tr - not determined (due to partial autoproteolysis affecting the result).

Protein denaturation - on the example of trypsin

To solution of Tr_CF_3 (400 μL, 400 μM in 100 mM NH_4OAc pH 8.0), solid DTT and urea were added (so that their final concentrations were 250 mM, 8 M respectively, note the dilution caused decrease in the protein concentration). After that, the mixture was incubated for 15 mins at 85 °C. Such a solution was diluted with D_2O (up to 10%) and used in ^{19}F NMR experiment. For proteins without disulfide bonds, DTT and thermal treatments were omitted.

Tryptic digest

For LC-MS/MS analysis: Standard reaction mixture was diluted with water (to a total volume of 140 μL) and buffer exchanged with a SpinTrap G25 according to the manufacturer's instructions (into 8 M urea and 50 mM NH_4HCO_3). DTT (2 μL, 200 mM in 50 mM NH_4HCO_3) was added to the solution of the modified protein (70 μL, 25 μg, ∼half of the desalted reaction mixture) and the mixture was incubated at 37 °C for 20 mins. After cooling to rt, a freshly prepared iodoacetamide solution (4 μL, 200 mM in 50 mM NH_4HCO_3) was added and the mixture was kept in the dark at rt for 20 mins. The excess iodoacetamide was quenched with the DTT solution (2 μL, 200 mM in 50 mM NH_4HCO_3). After dilution with the buffer (325 μL, 50 mM NH_4HCO_3), the protein was digested with porcine trypsin (4 μL, 0.25 μg/μL, 1:25 w/w to protein) in overnight incubation at 37 °C with 500 rpm shaking. The digestion was stopped by addition of aq. formic acid (430 μL, 0.2%) to reach a final protein concentration of ∼0.4 pmol/μL. A sample (10 μL) was desalted using ZipTip C18 according to the manufacturer's protocol and analysed by LC-MS/MS. In the

case of melittin, eqMb and PanC, the reduction and alkylation steps were omitted since those biomolecules do not contain Cys residues.

For eqMbCF$_3$ NMR analysis: To pooled samples of eqMbCF$_3$ after PD MidiTrap G-25 clean-up (∼16.5 mg, 50 mM NH_4HCO_3), porcine trypsin was added (0.66 mg, 1:25 w/w) and the reaction mixture was incubated overnight at 37 °C with 500 rpm shaking. The reaction mixture was analysed directly upon addition of D_2O (up to 10%). For melitting directly the crude reaction mixture was used for the digest.

Preparation of apomyoglobin

Apomyoglobin was prepared using an established methyl ethyl ketone extraction[250]. Briefly, solution of myoglobin (1%, ∼10 mg/mL) was acidified to pH ∼1.5 at 0 °C and extracted with an equal volume of cold methyl ethyl ketone. The aqueous layer was separated by pipetting out and extracted once more with half the volume of the organic solvent. The completion of heme removal was confirmed by measurement of UV-VIS spectra revealing the disappearance of the absorbance at 409 nm. The resulting, slightly straw-coloured, solution was buffer exchanged using a MidiTrap G25.

Structural and ligand binding PrOF experiments

The protein samples were diluted to final concentration of ∼ (100 – 300) μM in 100 mM NH_4OAc pH 8, containing 10% of D_2O unless stated otherwise. A concentrated stock of the ligand ((0.5 – 2) M) was prepared in the same buffer. In cases of sarcosine and trehalose, solid compounds were added to the protein solution. For determination of the berberine binding to lysozyme the general protocol for the affinity quantification

assay was adapted[223]. Such an approach was deemed valid as the assumptions of the assay were fulfilled (protein and the ligand being in a fast exchange), as per observation of a single, responsive resonance signal upon an addition of the compound.

Activity assay for lysozyme

A turbidimetric activity assay was conducted according to the modified general protocol adjusted for the 96-well plate reader[280]. The activity was determined upon a measurement of the decrease in an absorbance rate at 450 nm for cell suspension of *Micrococcus lysodeikticus* (0.02%, 240 μL) incubated with a lysozyme sample (20 μL). The following definition of unit was used: – one unit gives a ΔA450 of 0.001 per minute at pH 6.2 at 25 °C using a 0.02% suspension of *Micrococcus lysodeikticus* as the substrate in a 260 μL reaction mixture. The rate of lysis (activity) was determined by fitting of the linear region (first 2.5 minutes of the measurement) of a blank corrected absorbance decay. The experiment was performed twice and in duplicate and results averaged. Concentration of the samples was analysed with a micro-Bradford assay calibrated for the quantification of (1 – 10) μg/mL solutions of the protein. Ly_CF_3 was prepared according to the general protocol, and purified additionally by a SpinTrap G25 before the assay, according to manufacturer's instructions.

2.8 Appendix

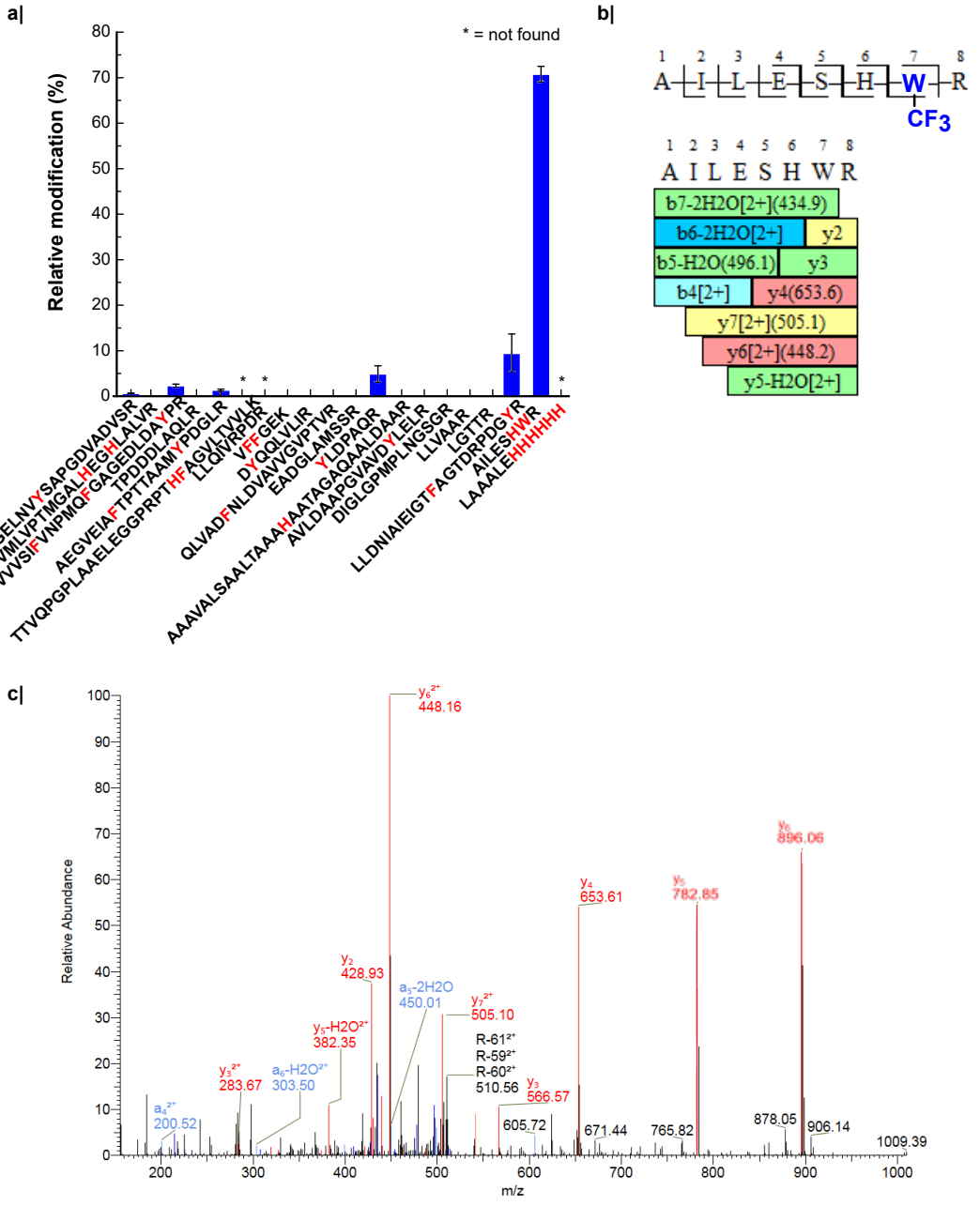


Fig. 2.29: The selectivity of trifluoromethylation for PanC was additionally confirmed with a tryptic digest of the modified protein and subsequent LC-MS/MS analysis ($n = 2$). **a**| The relative single CF_3 modification of different tryptic peptides. Results were averaged for two samples. **b-c**| Fragmentation spectra of he peptide with a CF_3-ylated Trp residue with annotated ions.

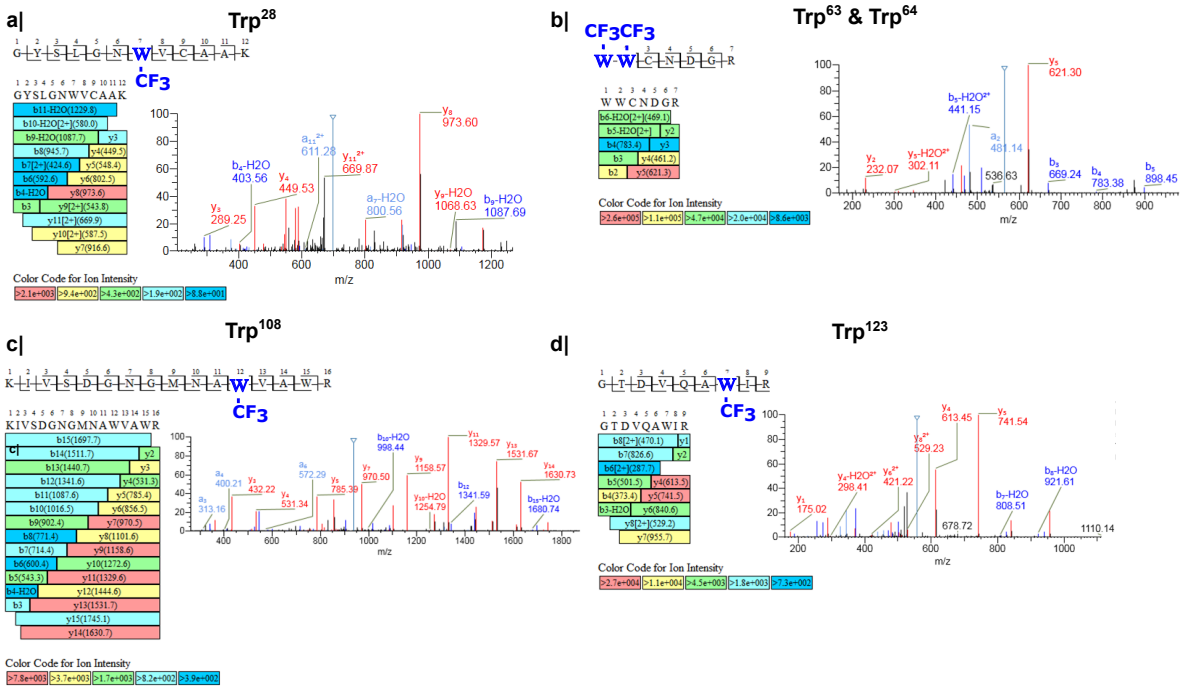
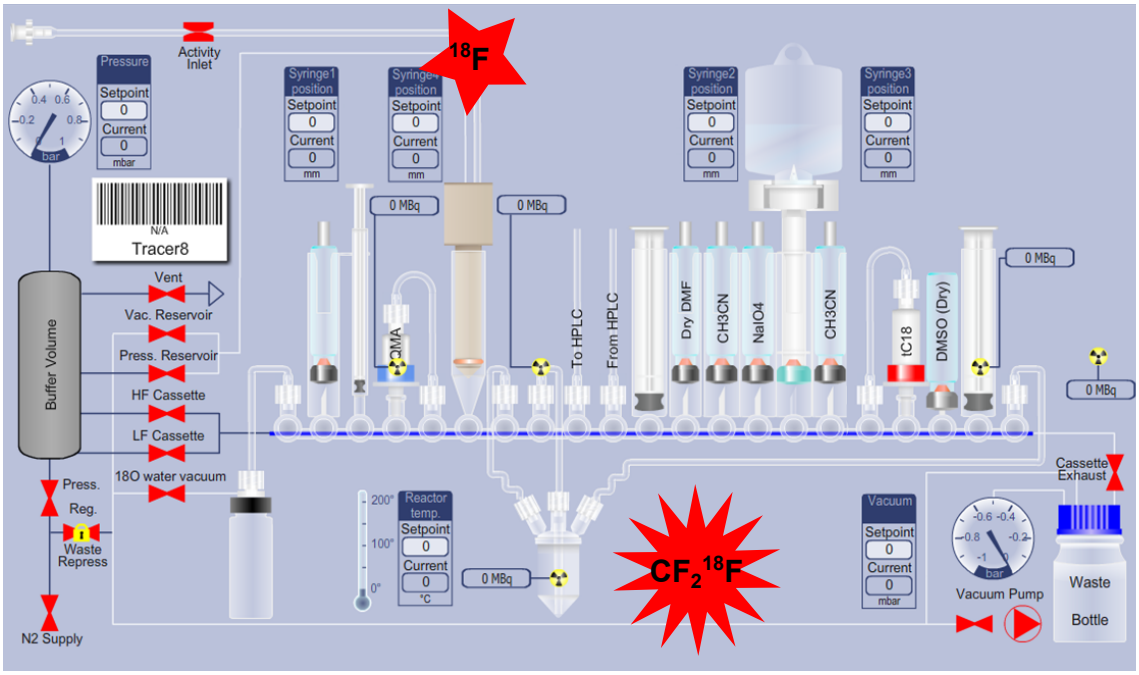


Fig. 2.30: LC-MS/MS analysis of digested Ly_CF_3 allowed identification of CF_3-ylated peptides. **a-d**| Peptide maps and representative tandem mass spectra for tryptic peptides modified at Trp. The fragmentation spectra corresponding to specific modification of Trp^{111} was not found, giving an overall coverage of 5/6 Trp residues.

Chapter 3

Towards the radiosynthesis of ^{18}F-trifluoromethanesulfinate and protein radiolabelling



Abstract

Protein radiofluorination is typically based on non-selective conjugation chemistry with nucleophilic side chains and relies on the introduction of bulky linkers commonly resulting in a significant alteration to the protein structure. The application of fluoroalkyl radical chemistry offers an alternative approach to the problem and Chapter 3 discusses the radiosynthetic preparation of a suitable radical precursor - ^{18}F-trifluoromethanesulfinate and subsequent protein radiolabelling attempts. Several synthetic routes based on the benzothiazole sulfone scaffold were evaluated using compounds with both ^{19}F and ^{18}F isotopes, allowing for a deeper understanding of the transformations. Ultimately, a novel method for the construction of the [^{18}F]-SCF_3 group was selected for optimisation. Although extensively investigated, only moderate improvements in yields were attained, leading to the necessity of automation of the radiosynthesis. Using a common commercial platform, the whole sequence could be translated for a synthesiser, and the desired radical precursor was obtained in satisfying quantities, allowing further studies. However, protein ^{18}F-radiotrifluoromethylation according to the proposed protocol was unsuccessful.

3.1 Introduction

An important aim of this thesis is to investigate the possibility of adaptation of the developed fluoroalkylation methodology for the other relevant isotope of fluorine - ^{18}F. Such a radiofluorination would provide an interesting alternative, enabling *direct* and potentially universal radiolabelling of proteins, without the need to introduce reactive prosthetic groups, therefore avoiding possible perturbations to the structure and function. At the start of this project, there were no reports of similar studies for either peptide or protein substrates.

To perform this investigation, the trifluoromethyl group in the radical precursor should have one of the fluorine atoms replaced with its radioactive isotope. A review of the literature (2017), revealed that the radiosynthesis of trifluoromethanesulfinate (TFMS) had not yet been accomplished. Therefore, the first task towards protein ^{18}F-trifluoromethylation would be to synthesise the labelled radical precursor using established radiosynthetic methodologies for the construction of the [^{18}F]CF_3 group.

An additional motivation to pursue this project stemmed from the versatility of the reagent. Trifluoromethanesulfinate, known since the 1950s as Langlois' reagent (sodium counterion), or more recently, Baran's reagent (zinc counterion), has been popularised as an excellent source of the electrophilic trifluoromethyl radical[281]. Since then, it has found use in numerous trifluoromethylation procedures for small molecules, under ever milder conditions[282]. This should not come as a surprise as the trifluoromethyl motif has been recognised to have desirable physicochemical properties by pharmaceutical and agrochemical industries and is now highly abundant among newly

developed synthetic drug molecules and derivatives[283]. Moreover, NaTFMS is characterised by a facile commercial availability and high operational simplicity, which altogether resulted in an research interest, as illustrated in Fig. 3.1 as a recent rapid growth in the number of publications and patents making use of it. At the time, it was surprising that there were no reports proposing its use in ^{18}F-trifluoromethylation, perhaps due to the challenging nature of radiosynthesis, which seemingly could be off-set by potential applications.

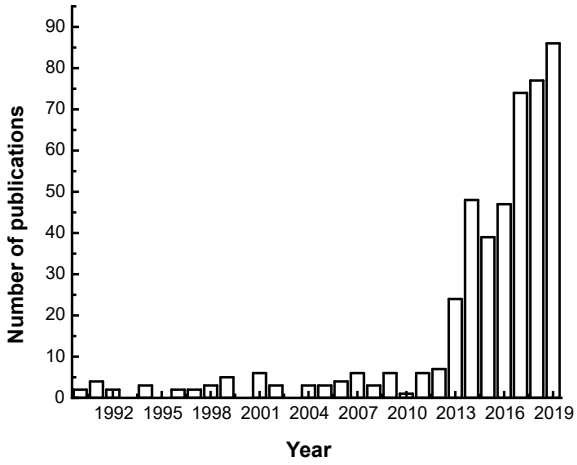


Fig. 3.1: Number of publications (including patents) using TFMS, traced from 1955 to 2019 (SciFinder - May 2020), reached a total of 497 (before 1990: 12 records).

3.1.1 $[^{18}F]CF_3$ versus $CF_3SO_2^-$ synthesis

As introduced in Chapter 1, ^{18}F is usually produced in nuclear reactions as an aqueous solution of ^{18}F-fluoride ($[^{18}F]F^-$) or much less commonly as gaseous $[^{18}F]F_2$. The latter is most commonly employed in electrophilic and/or radical fluorination either by direct use or by derivatisation into other, less reactive species[284]. However, use of $[^{18}F]F_2$ is connected with some serious drawbacks, such as the radioactivity content

limit, which result in the low specific activity (the measured activity per gram of material) of target compounds in addition to operational difficulties[285]. Therefore, a synthetic route based on an easier to handle, nucleophilic $[^{18}F]F^-$ could not only be more efficient but also may extend its applicability due to the straightforward production.

In general, there are several methods typically employed for construction of the ^{18}F trifluoromethyl group[286]:

a) $X \rightarrow {}^{18}F$ substitution in the CF_2X motif (X = leaving group),

b) $^{19}F \rightarrow {}^{18}F$ isotopic exchange in the CF_3 motif,

c) $^{18}F \rightarrow$ trapping of difluorocarbene ($:CF_2$) generated in situ, forming a nucleophilic $[^{18}F]CF_3^-$ species.

The most straightforward methods, (a), rely on nucleophilic substitution chemistry (S_N2), which often is moderately suited for the synthesis of the radiolabelled trifluoromethyl group most likely due to its stereoelectronic demands[287]. Nevertheless, due to the scarcity of other efficient methods, it often remains the default approach with a substrate dependent efficiency. The isotopic exchange method, (b), is explicitly limited in the specific activity, due to the presence of the competing ^{19}F. Finally, an interesting approach is to use a reactive difluorocarbene that can react with $[^{18}F]F^-$, (c), to form the trifluoromethylating reagent - $[^{18}F]CF_3^-$.

Defined as the measured activity per mole of compound, molar activity describes the ratio of radiolabelled (^{18}F-containing) material to its non-radioactive ^{19}F equivalent, and is an important characteristics of tracers, responsible for the quality of

PET image[288,289]. Radiotracers with high specific activity offer better signal-to-noise ratio (i.e. better image contrast) at saturating quantities, and as such are highly desirable. Unfortunately, all of the above methods for $[^{18}F]CF_3$ construction, result in radiolabelled products with relatively low specific activities (low $\sim$GBq/μmol), with particularly poor efficiency of the isotopic exchange method[290]. In some cases, the reasons for this are well understood i.e. from carrier addition or instability of the difluorocarbene intermediate. However, even for substitution chemistry, the obtained specific activity values are much lower than expected[286,291], highlighting challenging nature of such a radiosynthesis. For this reason the isotopic exchange pathway was not considered.

An extensive review of methods to prepare sulfinic acids and their salts (rich for aryl/aromatic but more limited for aliphatic ones), revealed two key pathways: direct sulfinylation with sulfur dioxide or elimination from a suitable sulfone precursor[292] - Fig. 3.2a. The former has been deemed difficult to realise in radiochemistry settings due to the gaseous state of the reagents - although a relevant precursor (e.g. $[^{18}F]$bromotrifluoromethane) was predicted to be easily accessed, its further reaction with SO_2 (or its equivalent in the form of sodium dithionite) would require high pressure and specialised equipment[293], not accessible in a typical ^{18}F-laboratory. On the other hand, the latter strategy exploits the fact that sulfinates are good leaving groups with a variety of alkyl and (hetero)aryl R substituents[294]. The cleavage of C-S bond is particularly fast and efficient for the benzothiazole scaffold treated with sodium borohydride[295]. Hence, it was envisaged that benzothiazole sulfone - **BzSOOCF$_3$**, would be a suitable precursor to the ^{18}F-labelled Langlois' reagent. A range of, above

discussed, standard radiolabelling methods were investigated in order to access this key intermediate - Fig. 3.2b.

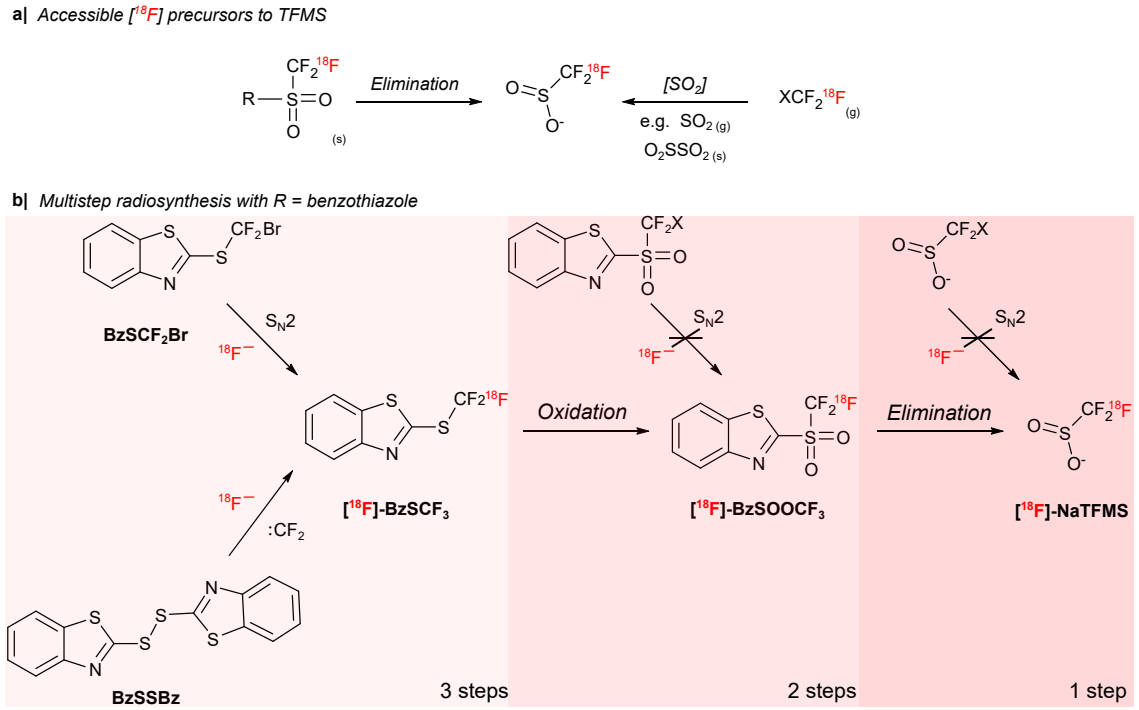


Fig. 3.2: **a**| Common synthetic strategies for the synthesis of trifluoromethanesulfinate. **b**| Analysis of synthetic routes to the key sulfone precursor and [^{18}**F**]-**TFMS**.

An ideal synthetic route has as few synthetic steps, involving the short-lived ^{18}F-isotope, as possible. Said that, the direct application of labelling methods (such as a or b) to the trifluoromethanesulfinate was not considered due to its charged nature and difficulties in accessing suitable precursors. Nevertheless, the radionuclide should be introduced as late as possible i.e. in this case the precursor would be a suitably substituted difluoroalkyl sulfone - $\mathbf{Bz(SOO)CF_2X}$. Such compounds are easily accessible[177] and seemingly a sulfone group should exert a favourable electron withdrawing effect facilitating nucleophilic substitution reaction, as is the case with carbonyl and cyano groups[296]. Unfortunately, α-halosulfonyl systems are notoriously unreactive

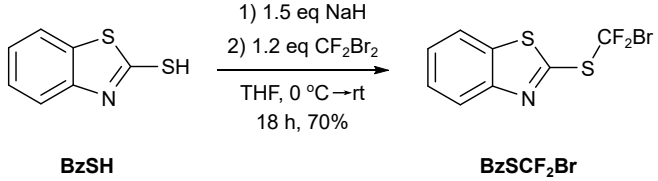
towards nucleophilic substitution, presumably due to the stereoelectronic hindrance at tetrahedral carbon[297]. For related reasons, difluorohalide sulfinate was also not considered as a viable precursor. However, the same problem is not encountered for aromatic sulfides[298], which then can be converted to the corresponding sulfones by a quick and facile oxidation step[299].

This chapter presents radiosynthesis of [^{18}**F**]**-TFMS** and its application to protein radiolabelling starting with ^{19}F feasibility studies. First, the reactivity of a suitable benzothiazole sulfide precursor in the nucleophilic substitution reaction is discussed, followed by a detailed investigation of the difluorocarbene benzothiazole disulfide trifluoromethylation and verification of the oxidation and elimination steps. Next section describes how these conditions were then adapted to ^{18}F-labelling, including optimisation of the [^{18}F]Cu-CF_3 mediated reaction. Following, the details of the automation of the full radiosynthesis sequence and molar activity determination are given. Finally, protein radiolabelling using [^{18}**F**]**-TFMS** will be discussed.

3.2 ^{19}F feasibility studies

3.2.1 $BzSCF_2Br$ as a substrate for nucleophilic displacement

First, a suitable precursor was synthesised using nucleophilic displacement of CF_2Br_2 to give **$BzSCF_2Br$** in a good yield - Sch. 3.1.



Sch. 3.1: Synthesis of the **$BzSCF_2Br$** precursor (10 mmol).

In the first step, nucleophilic fluorination of this precursor in a polar solvent was studied (KF, MeCN up to 90 °C - data not shown). However, ^{19}F NMR analysis of the reaction mixture revealed no formation of the expected product[177], with only the unreacted starting material present. Next, the addition of silver(I) salts was tested, which is known to mediate the halex (*halide exchange*) reaction for this type of substrates[298]. While addition of $AgNO_3$ resulted in no product formation, in the presence of AgOTf, **BzSCF$_2$Br** was converted to a new unexpected compound, as indicated by the observation of precipitate and ^{19}F NMR analysis - Fig. 3.3.

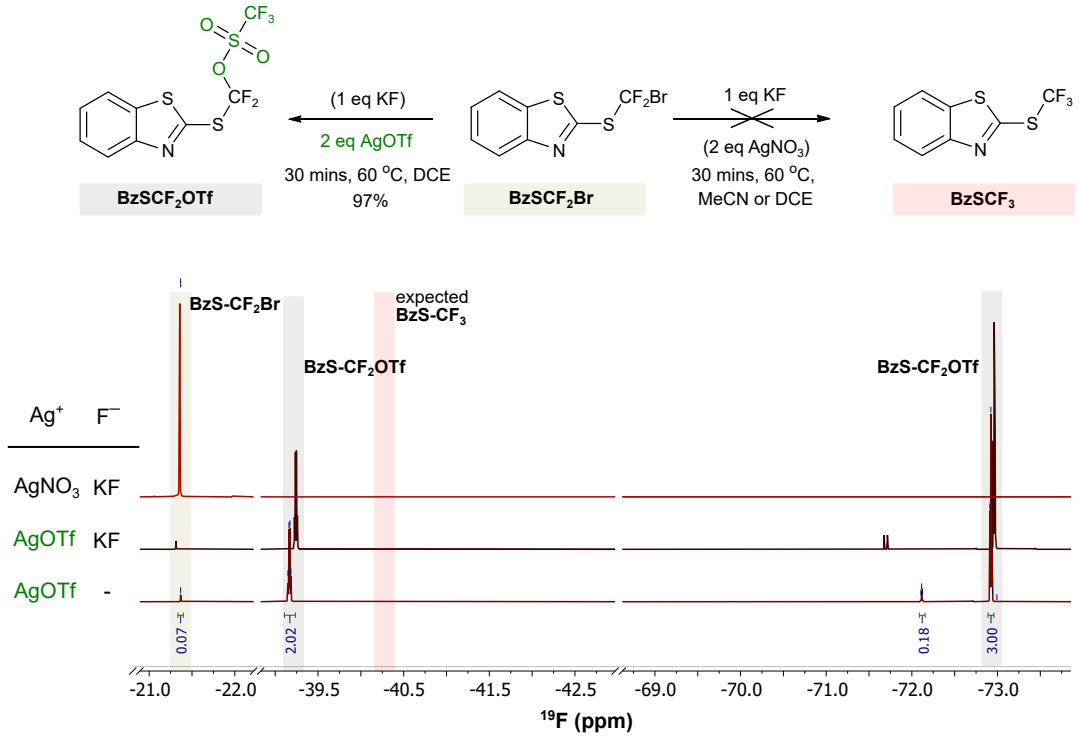


Fig. 3.3: ^{19}F NMR analysis of the crude reaction mixtures indicated that in the presence of AgOTf a new species is formed with a characteristic signal pattern (-39.2 (q, $J = 5.3$ Hz, 2F), -72.9 (t, $J = 5.3$ Hz, 3F)) elucidated to originate from the triflate derivative - **BzSCF$_2$OTf** with a structure proposed in the scheme.

Control experiments showed that this transformation still proceeded in the absence of

KF. Although unstable, it was characterised in a mixture with AgOTf in chloroform as a triflate derivative, which was further confirmed via GC-HRMS. While not common, such metathesis reactions have been observed in literature[300]. Considering that **BzSCF$_2$OTf** may be even better poised for the fluorination, further efforts were directed towards investigating its reactivity. A screen of solvents revealed an instability in polar solvents (DMSO, MeCN, MeOH, acetone, DMF) leading to a swift precipitation and complex mixtures, even in the absence of a F^- source. Limited to a less polar environment, the low solubility of the fluoride was suspected to inhibit the reaction. In order to aid it, a phase transfer catalyst (18-crown-6) was added and the reaction repeated. ^{19}F NMR, indeed reported a new transformation happening, however, with scarce amounts of the desired product formation - Fig. 3.4. Instead, the new species was hypothesised to originate from an attack of the F^- on the triflate sulfur, giving CF_3SO_2F that could be identified based on its characteristic ^{19}F signal pattern[301,302]. This type of reactivity is commonly observed for a related compound trifluoromethyl triflate[303] ($CF_3OSO_2CF_3$), for which the treatment with an anhydrous fluoride leads to release of OCF_3^- equivalent and CF_3SO_2F. It is postulated that the same type of reactivity is being observed here. Investigating other reaction parameters, such as temperature or solvents, did not favour the formation of **BzSCF$_3$**, and no other major fluorinated species were observed by ^{19}F NMR, indicating the decomposition of the benzothiazole fluoroalkyl leaving group.

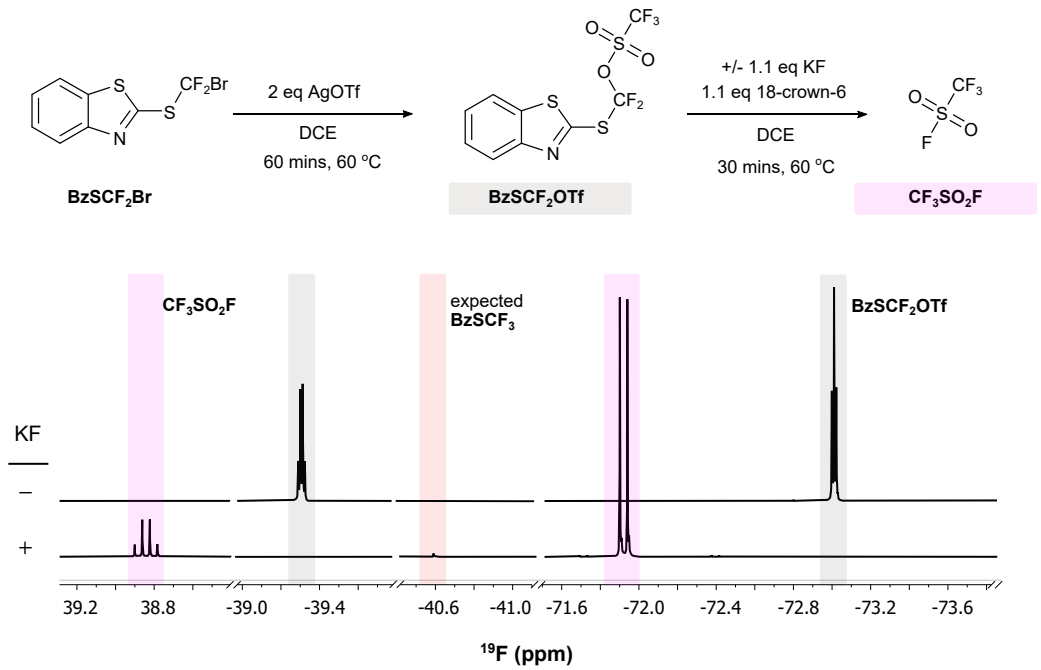


Fig. 3.4: ^{19}F NMR analysis of **BzSCF$_2$OTf** crude reaction mixtures indicated that in the presence of KF and a phase transfer catalyst (18-crown-6) a new species is formed with a characteristic signal pattern (38.85 (q, J = 18.3 Hz, 1F), −71.90 (d, J = 18.3 Hz, 3F)) hypothesised to originate from the CF_3SO_2F.

Reversing the reactivity i.e. achieving the desired transformation (cleavage of the C-O bond), has been reported to be difficult and the required conditions remain poorly understood, likely involving the need for HF, not readily available in the radiochemistry setup[304]. This led to the conclusion that Ag(I) mediation was not a suitable synthetic strategy for radiolabelling.

Finally, nucleophilic fluorination of **BzSCF$_2$Br** was carried out under conditions similar to those used in radiochemistry settings (short reaction times <20 mins, use of PTC and potassium carbonate)[305]. ^{19}F NMR revealed a minimal conversion (<1%), suggesting that although possible, such a reaction might not be an optimal way for labelling. Further optimisation was not performed, bearing in mind that general

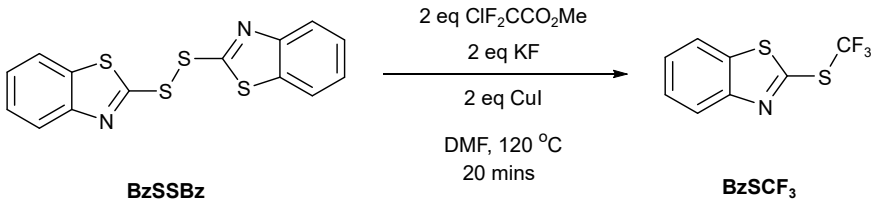
screening would be more advantageous during the radiolabelling step.

3.2.2 :CF_2 mediated S-S trifluoromethylation

Unsuccessful attempts at the classical nucleophilic substitution strategy prompted pursuing an alternative strategy based on the difluorocarboene precursor. Disulfides have been reported to be a suitable precursor to obtain the [^{18}F]SCF_3 motif in a reaction with [^{18}F]-fluoroform and a base[306]. This has confirmed that CF_3^- can easily react with a disulfide substrate as a partner electrophile[307], even in radiochemistry settings. Considering technical and regulatory difficulties in distilling [^{18}F]HCF_3, it was wondered whether a similar transformation could be achieved using an operationally simpler difluorocarbene derived trifluoromethylating equivalent. The capture of [^{18}F]F^- by :CF_2 has been well established, especially in the stabilising presence of Cu(I)[308]. However, no conditions for -S-S- trifluoromethylation using this compound have been reported in literature, with only scarce reports on a similar transformation - an oxidative addition of fluoroalkylated Cu(I) species, derived from organozinc reagents, into the S-S bond[309,310].

Starting from the previously established conditions[311], ^{19}F NMR analysis of a test reaction mixture confirmed the formation of the **BzSCF$_3$** product albeit in a very low yield. A series of control experiments were performed to determine the role of reaction components - Tab. 3.1. No product was observed in absence of CuI or when mercaptobenzothiazole (BzSH) was used as the substrate. Interestingly, traces of the product were observed also in the absence of KF, suggesting a participation of fluoride derived from :CF_2 degradation[286]. Next, the reaction was upscaled and products

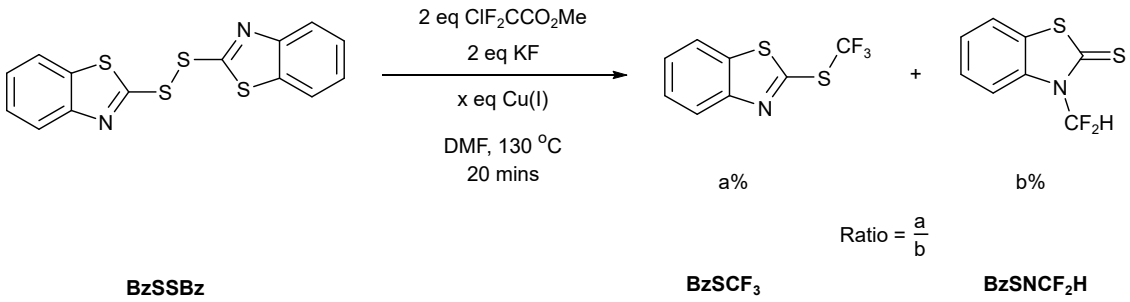
isolated and characterised. This allowed the recovery of another prominent fluorinated species elucidated to be the *N*-difluoromethylated derivative (**BzSNCF$_2$H**), typically obtained for thiazoles during a reaction with difluorocarbene[312].



Tab. 3.1: Control experiments of **BzSSBz** (2,2'-dithiobisbenzothiazole) trifluoromethylation confirmed the crucial role of difluorocarbene derived $CuCF_3$ in the product formation. Conversions were determined using quantitative ^{19}F NMR analysis.

Entry	**Substrate**	**:CF$_2$**	**F$^-$**	**Cu(I)**	**BzSCF$_3$**
1	BzSSBz	ClF_2CCO_2Me	KF	CuI	2%
2	BzSH	ClF_2CCO_2Me	KF	CuI	–
3	BzSSBz	$BrF_2CP(O)(OEt)_2$	KF	CuI	–
4	BzSSBz	ClF_2CCO_2Me	-	CuI	<1%
5	BzSSBz	ClF_2CCO_2Me	KF	-	–
6	BzSSBz	ClF_2CCO_2Me	KF	$AgNO_3$	–

Conditions were then screened to improve conversion and selectivity towards **BzSCF$_3$** - Tab. 3.2. It was found that employing a known Cu(I) ligand - TMEDA (*N*,*N*',*N*'-tetramethylethylenediamine)[308], led to an increase in both reactivity and selectivity. Varying the catalyst loading allowed for a further significant improvement, although changing of Cu(I) source (with or without TMEDA) led to similar or inferior results (Tab. 3.2 entries 4-8). Different solvents were not screened as DMF is known to play a crucial role in the reaction mechanism, stabilising the transient trifluoromethyl anion[313].

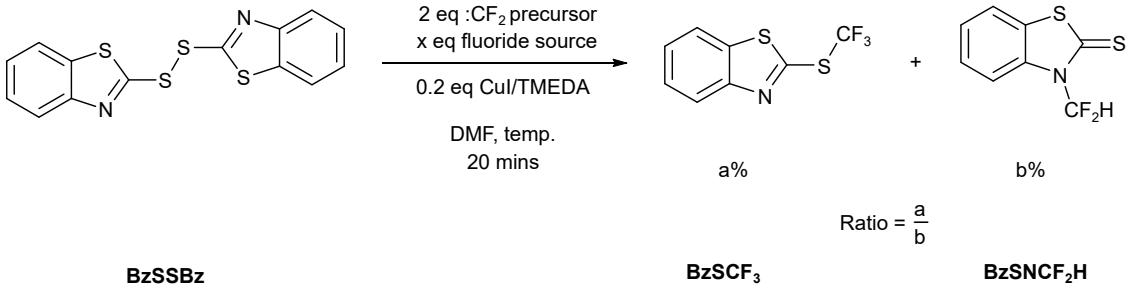


Tab. 3.2: Screening of different Cu(I) sources revealed the beneficial effect of the ligand TMEDA under substoichiometric loading. Column 'Ratio' refers to ratio between formed **BzSCF$_3$** and **BzSNCF$_2$H** products.

Entry	**Cu(I)/ligand**	**BzSCF$_3$**	**Ratio**
1	2 eq CuI/TMEDA	4%	0.5
2	1 eq CuI/TMEDA	5%	2
3	0.2 eq CuI/TMEDA	24%	3
4	0.2 eq $Cu_2(TMEDA)_2Cl_2(OH)_2$	11%	2
5	0.2 eq CuBr	11%	1.5
6	0.2 eq CuBr/TMEDA	14%	2
7	0.2 eq CuCl	9%	2
8	0.2 eq $Cu(ACN)_4OTf$	20%	2

A further improvement was brought about by the change to a more soluble fluoride source - CsF and an increase in its excess, clearly indicating the availability of F^- as the crucial factor. Although high temperature is typically employed to facilitate difluorocarbene formation, an excessive heating had a detrimental effect (Tab. 3.3 entry 3). Finally, similar :CF_2 sources did not perform well (presumably due to the low solubility of ClF_2CCO_2Na and an incomplete decomposition of ClF_2CCO_2Et as indicated by an intense starting material fluorine peak in ^{19}F NMR). Further optimisation was not pursued at this point postponing it to the relevant ^{18}F-trifluoromethylation. Low conversions for **BzSSBz** as well as diselenide (PhSeSePh[314] - 4% under conditions from Tab. 3.2 - entry 3), indicate that this methodology most likely did not

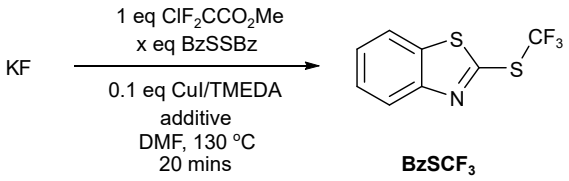
constitute a competitive ^{19}F-trifluoromethylation protocol[196].



Tab. 3.3: **BzSSBz** trifluoromethylation with variable temperature and CsF amounts. The best, semi-optimised conditions are highlighted in green.

Entry	**:CF_2**	**Temp. [°C]**	**F^-**	**$BzSCF_3$**	**Ratio**
1	ClF_2CCO_2Me	130	2 eq CsF	35%	5
2	ClF_2CCO_2Me	130	5 eq CsF	57%	22
3	ClF_2CCO_2Me	150	5 eq CsF	38%	5
4	ClF_2CCO_2Et	130	5 eq CsF	2%	2
5	ClF_2CCO_2Na	130	5 eq CsF	4%	0.5

As such, further tests were conducted in order to determine the performance under the conditions where fluoride is the limiting reagent and in the presence of common additives in ^{18}F-radiochemistry: potassium carbonate and a phase transfer catalyst - Tab. 3.4. Neither of these species, nor the change of the solvent to MeCN, negatively affected the reaction and overall yielded similar conversions. Although the efficiency dropped following the 10 fold decrease in scale, the product was still formed, raising hopes of success in ^{18}F-settings. To complicate the prediction it is unknown what fraction of fluoride was soluble under these conditions. However, employing a large excess of the difluorocarbene source, although unfavourable for the specific activity, should ensure that all of the fluoride would be consumed in the construction of the trifluoromethyl group.



Tab. 3.4: **BzSSBz** trifluoromethylation treating KF as the limiting reagent as well as in the presence of typical additives in radiochemistry. All calculations with respect to fluoride.

Entry	BzSSBz	Scale	Additive	F^- conversion
1	0.5 eq	300 μmol	–	12%
2	0.5 eq	300 μmol	MeCN[a]	13%
3	0.5 eq	300 μmol	K_2CO_3[b]	15%
4	1 eq	300 μmol	–	(10±3)[d]%
5	1 eq	300 μmol	MeCN[a], K_2CO_3[b], PTC[c]	13%
6	0.5 eq	30 μmol	–	(3.3±0.7)[d]%
7	1 eq	30 μmol	–	4%

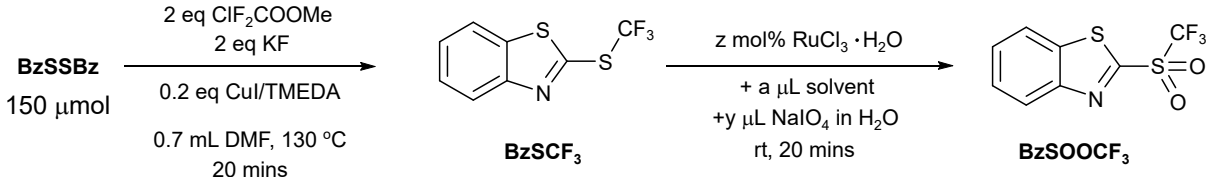
[a] - 6 eq, [b] - 0.75 eq, [c] - phase transfer catalyst dicyclohexano-18-crown-6, 1 eq
[d] - average from $n = 3$ reactions.

3.2.3 Verification of the oxidation-elimination sequence

Having investigated the synthetic route to **BzSCF$_3$**, the attention was turned to study the actual sequence of reactions leading to the desired trifluoromethanesulfinate. First, oxidation to the corresponding sulfone was performed using an established procedure[177], with ruthenium(III) chloride as a catalyst and sodium periodate as the oxidant. Although some of the reported procedures[315,316] have made use of the addition of chlorinated co-solvents (CCl_4, DCM), it was found out that a mixture of MeCN and H_2O worked equally well for this substrate. Under the high excess of the periodate, the reaction was completed in under 1 h as indicated by ^{19}F NMR analysis.

A practical consideration when working with ^{18}F would be the ability to perform a one-pot oxidation on the crude reaction mixture containing the **BzSCF$_3$** substrate

to reduce the total time of radiosynthesis. Initial oxidations on a small scale without DMF removal revealed no product formation by ^{19}F NMR analysis - Tab. 3.5.



Tab. 3.5: **BzSCF$_3$** oxidation sequence using crude reaction mixtures under different conditions and solvent composition. Amounts of oxidising agents are calculated with respect to **BzSCF$_3$** in the crude reaction mixture (∼25% from **BzSSBz**). **BzSOOCF$_3$** conversion was determined using quantitative ^{19}F NMR analysis versus **BzSCF$_3$**.

Entry	**Crude mix.**	**Solvents** [μL]		**Ox. agents** [eq]		**Conversion**
	DMF [μL]	H_2O	Organic	$RuCl_3$	$NaIO_4$	**BzSOOCF$_3$**
1	350	350	–	0.5	7.5	–
2	350	350	350 $CDCl_3$	0.5	7.5	–
3	700	700	350 MeCN	0.5	7.5	–
4	350	700	3000 MeCN	0.5	15	–
5	350	3000	–	0.5	10	–
6	300	350	50 MeCN	0.6	8	–
7	200	350	100 MeCN	0.9	13	2%
8	100	350	200 MeCN	1.7	25	96%
9	50	350	300 MeCN	3.4	51	100%
10	50	350	300 DMF	3.4	51	100%
11	25	350	325 MeCN	7	102	100%

While the addition of different co-solvents did not improve the conversion, employing a large excess of both the catalyst and oxidant resulted in full conversion to the desired product **BzSOOCF$_3$** in 20 mins. This suggested that direct and fast oxidation of **BzSCF$_3$** in the crude reaction mixture might be achievable in radiochemistry settings.

The final set of experiments was aimed to verify the elimination of the desired sulfinate. Expectedly, the reaction on the crude mixture was highly exothermic and

gave minimal conversions therefore this way was not deemed feasible. In order to isolate **BzSOOCF$_3$**, the crude solution obtained under optimised conditions (Fig. 3.5a) was diluted with H_2O to form a suspension and passed through a C18 cartridge (pre-activated with MeCN). It was found that the water addition was a necessary step for the complete adsorption of the compound. In this way, no **BzSOOCF$_3$** was found in the flowthrough as verified by ^{19}F NMR analysis. The elution from the cartridge was straightforward with low amounts of EtOH (2 mL) after washing the cartridge with 2 mL aqueous 10% EtOH solution.

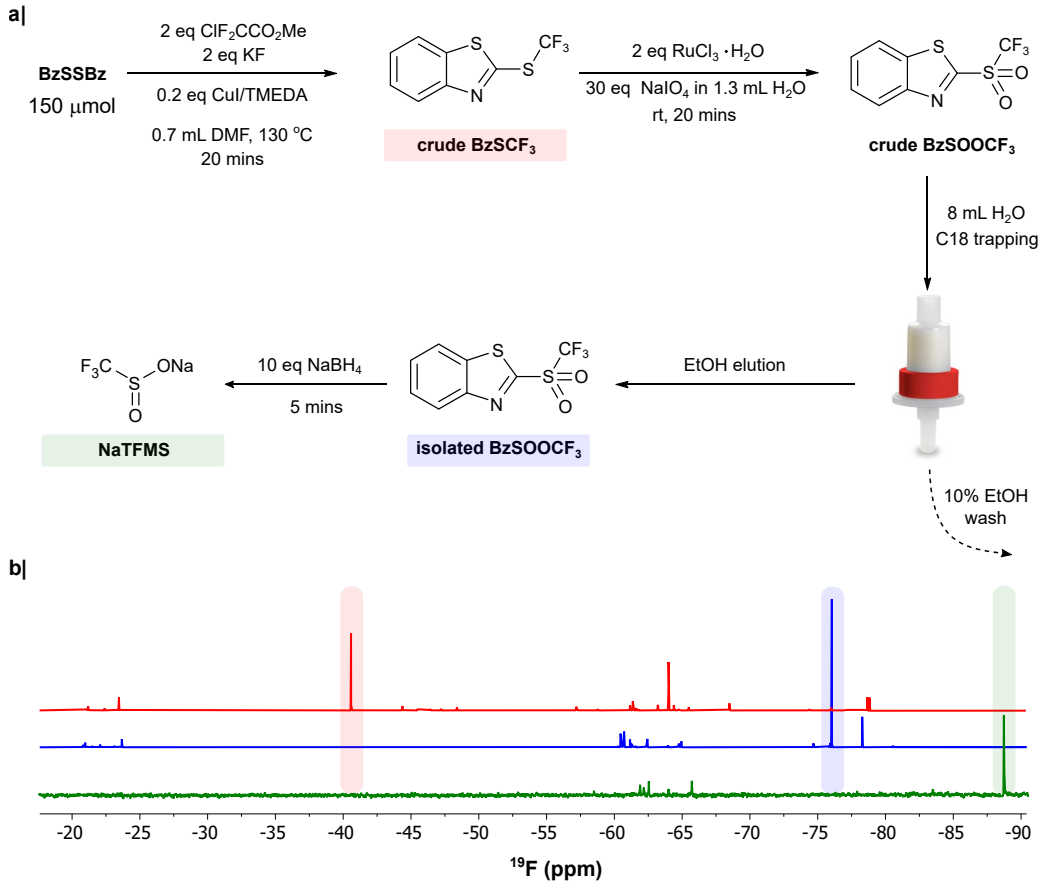


Fig. 3.5: **a**| ^{19}F feasibility study for the formation of NaTFMS via the :CF_2 mediated synthetic route. The trial did not include isolation of the product. **b**| ^{19}F NMR traces of reaction mixtures following each transformation with highlighted species of interest.

Addition of an excess of sodium borohydride resulted in a rapid (< 5 mins) elimination of the desired sulfinate, as confirmed by ^{19}F NMR. Gratifyingly, the final solution predominantly contained the desired sulfinate - Fig. 3.5b. Semi preparative HPLC would be used to isolate the pure ^{18}F-product. Together these results suggested that this methodology may be feasible for radiosynthesis of NaTFMS, which was the object of further investigation.

3.3 ^{18}F radiochemistry

3.3.1 Radiolabelling of the radical precursor

Two principal methods of screening were employed:

1. Fully automated radiosynthesis in a FastLab^{TM} module synthesiser (GE Healthcare), including the preparation of $[^{18}F]F^-$, the addition of reagents and radiolabelling reaction. All analyses were performed manually. The operations were controlled remotely via dedicated software.

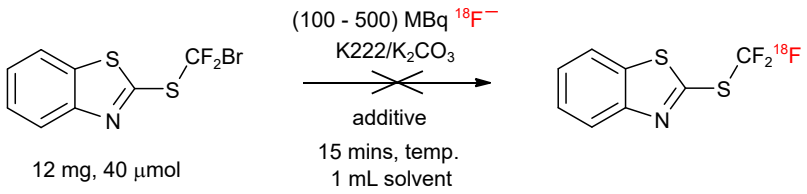
2. Manual radiosynthesis using $[^{18}F]F^-$ aliquots, prepared beforehand in the synthesiser as a solution in relevant solvents.

Such an approach ensured that a higher number of conditions could be screened with a different level of starting radioactivity. For all experiments, $[^{18}F]F^-$ was prepared following standard methods, by trapping from the target water on an anion exchange cartridge and subsequent elution with a solution of a base (here, potassium carbonate) and a metal-chelating cryptand - Kryptofix^{TM} 222. The reactive $[^{18}F]F^-$ was obtained after the final azeotropical drying and dissolution in a desired solvent[317].

This preparation is typically employed in radiosyntheses of most popular ^{18}F-tracers and, in addition to the high reproducibility, it also ensures its applicability throughout different laboratories.

To comprehensively evaluate the outcome of reactions, several analysis methods were used. The crude reaction mixtures were assayed with radio-TLC (MeOH) that was used to determine the conversion to radiolabelled compounds, allowing convenient separation from unreacted inorganic [^{18}F]F^{-}. The purity was then further evaluated with a radio-UPLC analysis (in which unreacted [^{18}F]HF was retained on the column due to the use of acidified mobile phases)[318], which also reported on the side products distribution and confirmed the identity of the product (retention time comparison with the reference). The mathematical product of numbers: conversion from radio-TLC (MeOH) and purity from radio-UPLC is reported as the radiochemical conversion (RCC). An additional test was the radio-TLC analysis using $CHCl_3$ that was used to validate the results based on a known position of **BzSCF$_3$** (Rf $\sim$0.7).

Initial trials were aimed at verification of low reactivity of **BzSCF$_2$Br** in nucleophilic radiofluorination - Tab. 3.6.



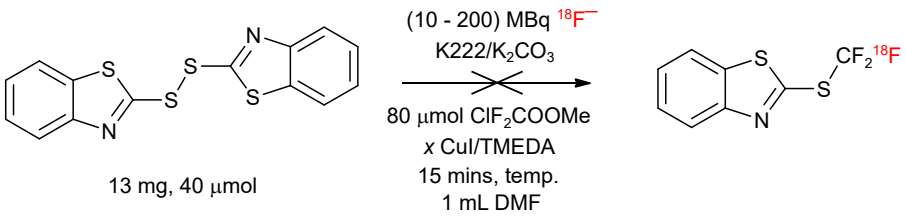
Tab. 3.6: **BzSCF$_2$Br** radiolabelling under different conditions. DMA = dimethylacetamide, NMP = *N*-Methyl-2-Pyrrolidone.

Entry	**Additive**	**Temp** [°C]	**Solvent**	**TLC**	**UPLC**	**RCC**
1[a]	–	120	MeCN	<2%	0%	–
2	–	145	DMF	<2%	0%	–
3	Pyridine[b]	145	DMF	<2%	0%	–
4	$AgNO_3$[c]	60	DCE	3%	0%	–
5	AgOTf[c]	125	DMF	4%	0%	–
6	AgOTf[c]	80	MeCN	<2%	0%	–
7	AgOTf[c]	125	NMP	<2%	0%	–
8	AgOTf[c]	125	DMA	<2%	0%	–
9[a,d]	AgOTf[c]	60	DCE	(3 ± 1)%	0%	–

[a] - automated radiosynthesis, [b] - 150 μmol, [c] - 80 μmol, [d] - $(n = 2)$

Indeed, in all of those labelling experiments only a minimal amount of radiolabelled organic compounds was obtained, among which the desired product was not found. These results were in agreement with the feasibility studies, and no further trials were conducted apart from the shown screening of solvents and silver additives.

Instead, the more promising :CF_2 mediated strategy was pursued. However, the first **BzSSBz** radiolabelling experiments using the optimised conditions from the feasibility study failed - Tab. 3.7.



Tab. 3.7: **BzSSBz** radiolabelling trials under different conditions.

Entry	**TMEDA&CuI** [μmol]	**Temp.** [°C]	**TLC**	**UPLC**	**RCC**
1[a]	8	120	<1%	–	–
2[b]	8	150	7%	–	–
3	8[c]	150	8%	–	–
4	8	150	11%	–	–
5	80	145	13%	–	–
6	0.8	145	13%	–	–
7[d]	80[e]	145	(13 ± 2)%	(8 ± 5)%	(1 ± 1)%

[a] - automated radiosynthesis with 2.5 mg of BzSSBz,

[b] - different reagents order, [c] - freshly distilled TMEDA

[d] - $n = 2$, [e] - CuI from a different supplier

BzSSBz has a limited solubility in DMF (∼2.5 mg/mL) which limited the amount of precursor that could be used with the synthesiser. Therefore, further screenings were performed manually. Repeating the reaction with different amounts of the CuI/TMEDA catalyst did not change the outcome, nor changes in the reagent order addition and purification of the amine ligand. To exclude the possibility for the excessive formation of radiolabelled volatiles (e.g. ^{18}F-fluoroform)[319], the activity was measured at each stage of experimental procedures - Tab. 3.8. Such an analysis, upon decay-correction, should yield the total starting activity if there are no gaseous losses. It was found that the radioactivity escape was < 10%, indicating that such off-side reactivity was not a likely cause of the poor reaction performance.

Tab. 3.8: Analysis of the radioactivity distribution during **BzSSBz** trifluoromethylation under conditions from Tab. 3.7- entry 7. DCA = decay corrected activity. w/rxn mix = with reaction mixture. QMA = $[^{18}F]F^-$ trapping cartridge.

Item	**Activity**	**Elapsed**	**DCA**	
	[MBq]	[mins]	[MBq]	[%]
Starting $[^{18}F]F^-$	390	0	390	
Residual	0.6	0	0.6	
Used for the synthesis			389.4	100
QMA	2.9	42	4	1.0
Target vial	1.9	42	3	0.8
Reactor w/rxn mix	265.4	43	348	89.4
Other elements	2	48	3	0.8
		Loss	**31**	**8.0**

Traces of the radioactive **$[^{18}F]$-BzSCF$_3$** product were observed when CuI from a different supplier was used, which suggested that the quality of the metal salt was an important factor - Fig. 3.6a. The product peak could also be observed in the UV chromatogram - Fig. 3.6b. To further confirm its identity, the reaction mixture was spiked with the ^{19}F reference, which caused an increase of absorbance at the same retention time - Fig. 3.6c.

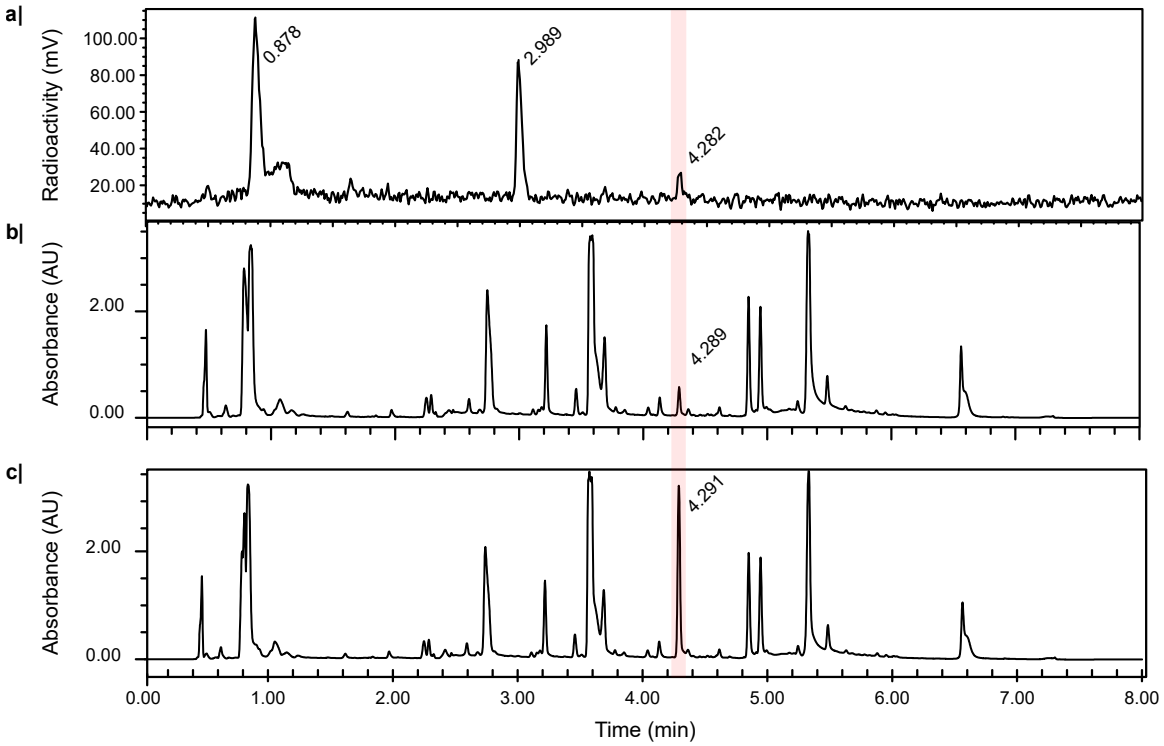
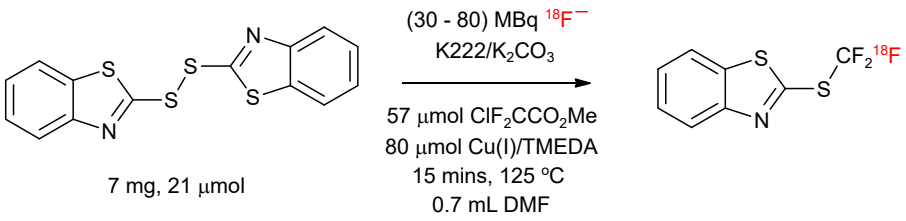


Fig. 3.6: **a**| Radio-chromatogram and **b**| UV-chromatogram of the reaction mixture in which traces of the expected product (highlighted in red) were identified (Tab. 3.7 - entry 7) **c**| UV-chromatogram of the reaction mixture spiked with $\mathbf{BzSCF_3}$ reference.

Following this observation and considering the fact most of the Cu(I) sources were almost equally efficient in the feasibility study, different salts were investigated - Tab. 3.9. The screening revealed a potential superior reactivity of CuCN and CuBr, with the former producing less reproducible results as well as higher number of side products, leaving CuBr as the preferred source of Cu(I).



Tab. 3.9: **BzSSBz** radiolabelling with different Cu(I) salts.

Entry	Cu(I)	TLC	UPLC	RCC
1[a]	$CuBr{\cdot}S(CH_3)_2$	14%	–	–
2	$CuBr{\cdot}S(CH_3)_2$	10%	9%	1%
3	CuCl	15%	7%	1%
4[b]	CuCN	$(26 \pm 3)\%$	$(13 \pm 7)\%$	$(5 \pm 3)\%$
5[b]	CuBr	$(24 \pm 2)\%$	$(16 \pm 3)\%$	$(4 \pm 1)\%$

[a] - no TMEDA addition, [b] - $n = 3$

Further optimisation of the reaction stoichiometry and temperature led only to small changes (< 3%, of a low relevance as the reactions were not done in replicates). As such, an additional screen of additives (ligands) and solvents was performed under the standard conditions - Appendix (Tab. 3.11). Unfortunately, this did not improve the efficiency of the reaction and optimisation was stopped at this point, bearing in mind that using high level of the starting activity might allow for the generation of sufficient amounts of ^{18}F-labelled reagent to test protein radiolabelling. Therefore, the project was continued with investigation of the oxidation and elimination sequence as well as the possibility of automation of the radiosynthesis.

3.3.2 Automation and optimisation of the oxidation-elimination sequence

The manual radiosynthesis of [^{18}**F**]-**BzSCF$_3$** was adapted to automation with the synthesiser. Due to the low solubility of the **BzSSBz** and the technical inability to

transfer suspensions, the exact stoichiometry could not be replicated and amounts of reagents were reduced. This allowed isolation of [^{18}**F**]-**BzSCF**$_3$ after C18 trapping with (0.5 ± 0.2)% RCC (n = 2). To verify the oxidation-elimination procedure, the sequence was first performed manually following the automated synthesis of [^{18}**F**]-**BzSCF**$_3$. The compound was readily oxidised after 5 mins with an aqueous solution of the oxidising reagents ($RuCl_3$ & $NaIO_4$), forming quantitatively [^{18}**F**]-**BzSOOCF**$_3$ - Fig. 3.7a-b. The identity of the product was confirmed by spiking the reaction mixture with the reference material - Fig. 3.7c-d.

Next, the elimination of the corresponding sulfinate was investigated. The initial investigation focused on reactions of the trapped sulfone directly on the C18 cartridge to reduce the radiosynthesis time. However, passing a $NaBH_4$ solution either in EtOH or in a aqueous mixture did not allow the elimination process to proceed smoothly as a large amount of activity was trapped inside (both as [^{18}**F**]-**BzSOOCF**$_3$ and [^{18}**F**]-**NaTFMS**), even using excessive volumes of the reagent. Therefore, it was concluded that the elimination step must occur in solution. Indeed, it was found that the elution of the sulfone from the cartridge with MeCN followed by a treatment with an excess of $NaBH_4$ led to the complete conversion to the desired product - Fig. 3.8.

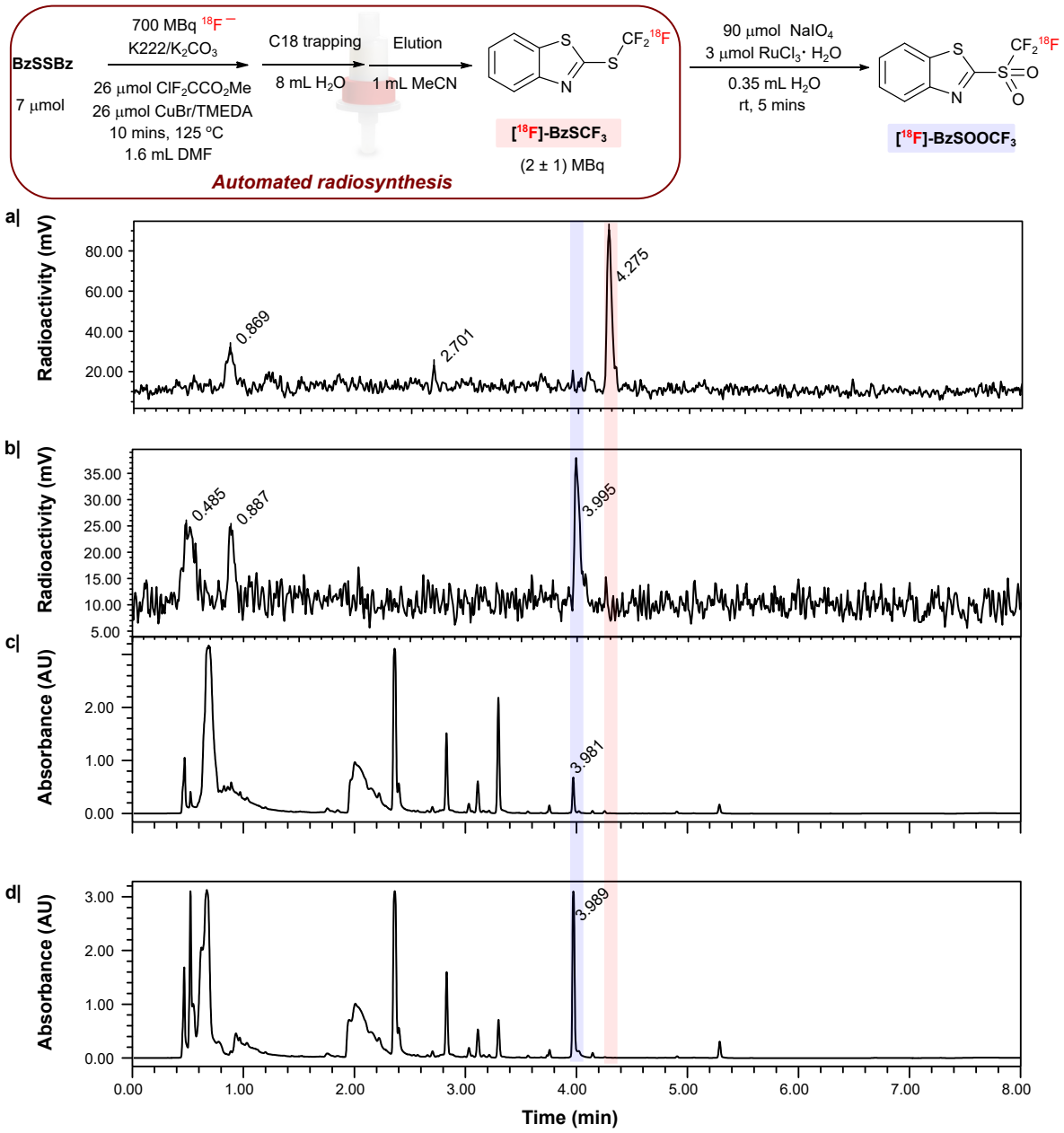


Fig. 3.7: Radio-chromatograms of: **a**| the [^{18}**F**]-**BzSCF**$_3$ solution eluted from C18 cartridge after automated synthesis, **b**| the [^{18}**F**]-**BzSOOCF**$_3$ crude reaction mixture after a manual oxidation, with the peaks highlighted in red and blue, respectively. UV-chromatograms of **c**| the oxidation reaction mixture and **d**| the mixture after the spiking experiment, where the identity of the desired product was confirmed by the addition of ^{19}F-reference. Reported activities are non-decay corrected.

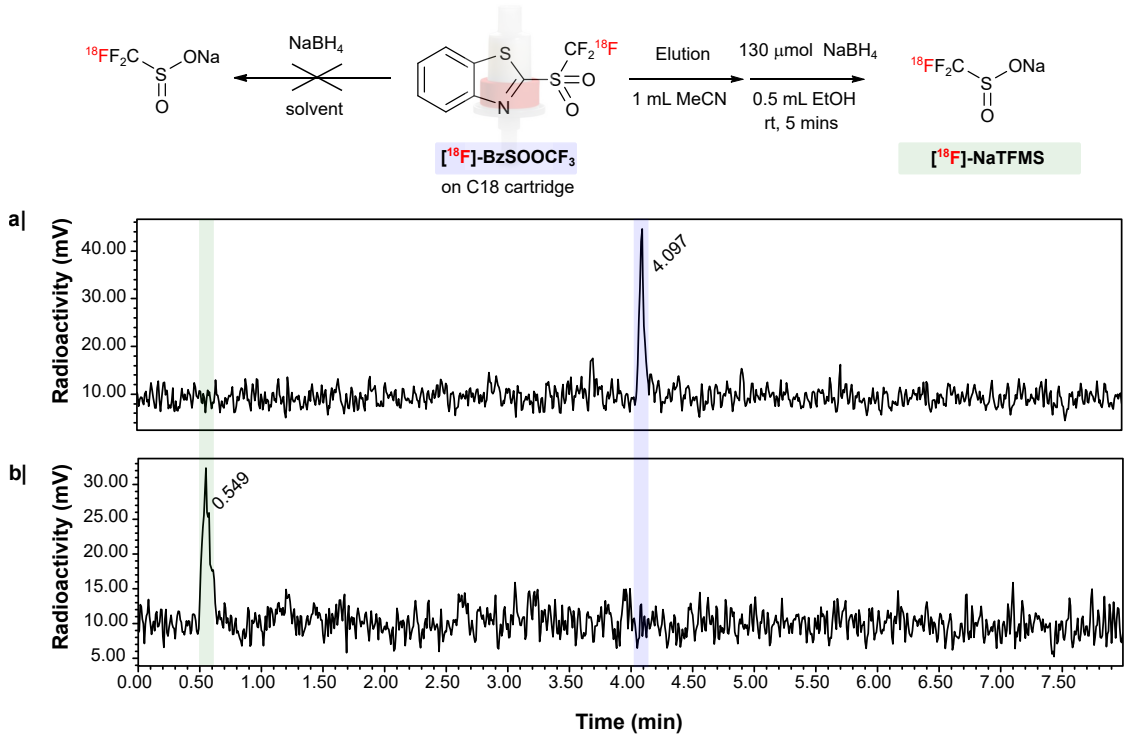
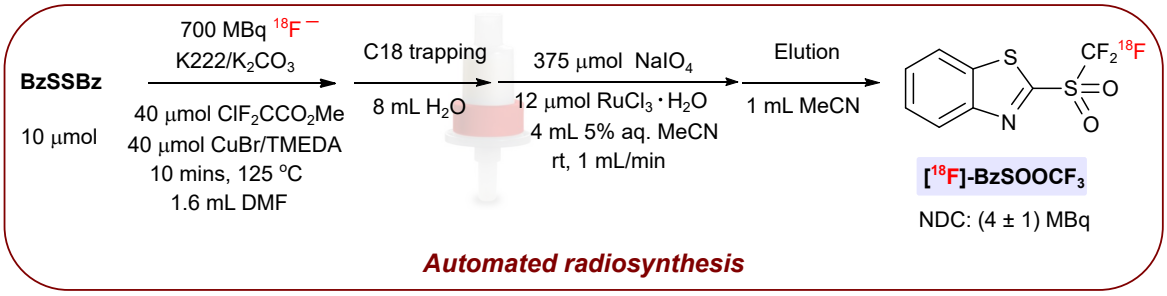


Fig. 3.8: Radio-chromatograms of **a**| the solution eluted from C18 cartridge and **b**| the reaction mixture after the elimination, with [^{18}**F**]-**BzSOOCF**$_3$ and [^{18}**F**]-**NaTFMS** peaks highlighted in blue and green, respectively.

In order to develop a reliable analytical method (based on RP-UPLC) for the detection of [^{18}**F**]-**TFMS**, it was necessary to use an additive of an ion pairing reagent (NBu_4Br) to better retain the compound on the column (retention time change from 0.6 min to 2.1 min). Following this strategy, a semi-preparative HPLC procedure was also developed using an isocratic elution with 20% aqueous MeCN and 5 mM NBu_4Br as the eluents.

Moving forward, the oxidation procedure was implemented into the automated sequence. Unfortunately, when aqueous the solution of oxidants was added straight to the crude reaction mixture, the resulting suspension could not be transferred through the casette onto a C18 cartridge without over-pressurising the system and causing

the tubing to burst. Inclusion of an additional filter and fortified transfer tubes only partially alleviated the problem, with only low volumes being passed before the over-pressure. However, this was not the case if just the crude radiolabelling reaction mixture was applied to the cartridge. Therefore, the oxidation on the C18 cartridge was carried out. Through optimisation of this procedure (not shown) it was found out that slow (1 mL/min) flowing of the aqueous (with 5% MeCN) solution of oxidants through the cartridge with the adsorbed [^{18}**F**]-**BzSCF$_3$** resulted in complete conversion to the sulfone. Subsequent elution with MeCN gave [^{18}**F**]-**BzSOOCF$_3$** in an overall RCC of (0.9±0.2)% ($n = 2$) for the two-step transformation - Sch. 3.2. A slight improvement in the conversion (compared with just [^{18}**F**]-**BzSCF$_3$** radiosynthesis) was attributed to the higher amounts of the precursor and reagents used.



Sch. 3.2: Automated radiosynthesis of [^{18}**F**]-**BzSOOCF$_3$**.

Finally, the full sequence was automated, allowing the production of [^{18}**F**]-**NaTFMS** in (1.0±0.2)% non decay corrected RCC ($n = 5$) with a decent radiochemical purity of (77±16)% - Fig. 3.10. The final reaction mixture was further purified by RP-HPLC to isolate the reagent. Afterwards, the HPLC eluent could be removed during the reagent concentration via trapping on the C18 cartridge and elution with a low volume of MeCN.

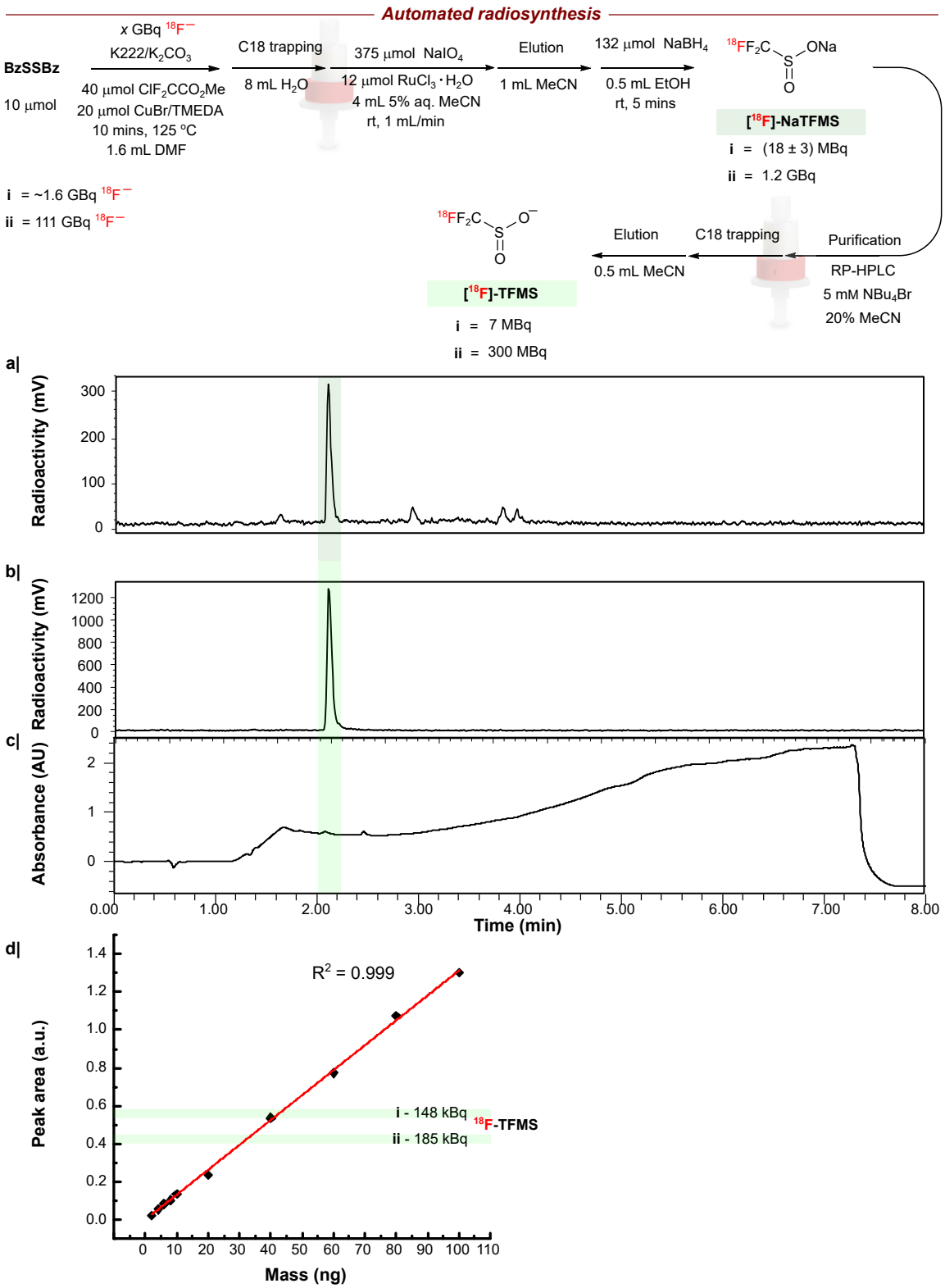
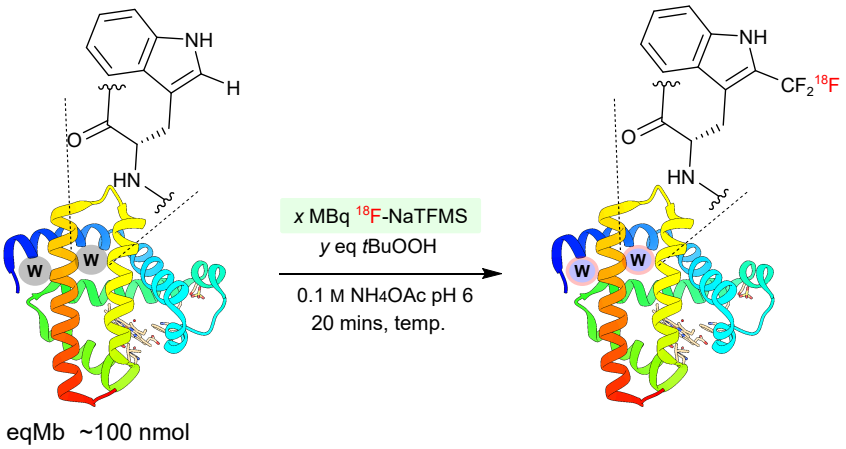


Fig. 3.9: Radio-chromatograms of **a**| the crude reaction mixture and **b**| the reaction mixture after purification with the [^{18}**F]-TFMS** peaks highlighted (dark and light green, respectively). The molar activity was determined based on **c**| UV-absorbance of the isolated product matched against **d**| the calibration curve prepared using [^{19}F]-NaTFMS.

The solution of ^{18}**F-TFMS** was then either directly diluted or evaporated under a stream of nitrogen to make up an aqueous stock solution that would be compatible for the subsequent protein radiolabelling. The total synthesis and purification time was ∼100 min, just below one ^{18}F half-life. In order to perform screening of protein radiolabelling, synthesis of the reagent starting with a very high activity was carried out (∼111 GBq). The automated procedure delivered the crude reagent as expected - Fig. 3.8a. Subsequent isolation yielded 300 MBq (∼0.3% NDC RCY) of the purified reagent - Fig. 3.8b. Gratifyingly, the molar activity improved from 0.5 GBq/μmol, when synthesised using 100x lower starting activity, to 1 GBq/μmol - Fig. 3.8c-d.

3.3.3 Protein radiolabelling

With the ^{18}F-reagent in hand, radiotrifluoromethylation of a model protein was investigated. Recreating the conditions from the ^{19}F -feasibility study - Fig. 2.24, using a relatively high concentration of myoglobin ($>$ 500 μM) and elevated temperatures, did not allow efficient formation of the radiolabelled protein even after prolonged reaction times - Tab. 3.10. Using a higher amount of the radical precursor, under forcing condition with respect to stability of biological systems, resulted in only a trace of plausible radiolabelled product being detected - Fig. 3.10a-b. Due to the operational constraints, further radiosyntheses of the reagent were not possible, hence different conditions were not screened. The Gouverneur group has developed a more efficient radiosynthesis of [^{18}**F**]**-TFMS**[320]. That strategy also utilised a :CF_2 mediated pathway but in a single step using *N*-methylmorpholine · SO_2 as a sulfur source, which led to the desired product with RCY of ∼6%.



Tab. 3.10: Trifluoromethylation of eqMb (~500 μM, 0.1 M NH_4OAc, 20 mins) with varying amounts of reagents and temperature. All calculations with respect to protein. Experiment from entry 7 was performed in collaboration with Prof V. Gouverneur group.

Entry	**^{18}F-NaTFMS**		***t*BuOOH**	**Temp.**	**RCC**
	[MBq]	[eq]	[eq]	[°C]	
1	3.5	0.1	2.5	rt	–
2	3.5	0.1	5.0	rt	–
3	7	0.2	2.5	rt	–
4	7	0.4	2.5	rt	–
5	7	0.2	25	40	–
6	19	0.5	2.5	40	5%
7[a]	5	5	10[b]	rt	–

[a] - experiment performed by the Gouverneur group

[b] - additional 2.5 eq $Fe_2(SO_4)_3$ as an additive

In a kind collaboration with Dr Choon Wee Kee and Dr Osman Tack radiolabelling of myoglobin, with [^{18}**F]-TFMS** obtained via the new method, was revisited. Unfortunately, also in this case no ^{18}F-products were observed. Interestingly, the subsequent LC-MS analysis of the decayed reaction mixture revealed a significant formation of eqMbCF_3 - Fig. 3.10d, which can be only explained by a relatively low molar activity of the radical precursor (0.01 GBq/μmol)[320]. As a consequence, the majority

of trifluoromethylation occurred via the [^{19}F]-NaTFMS (^{18}F-isotope bearing reagent constituted only ~$1.6{\cdot}10^{-5}$ % - based on the maximum theoretical molar activity[262]). An experiment under identical conditions but using [^{19}F]-NaTFMS gave similar results (upper spectrum - Fig. 3.10d) further reinforcing the hypothesis.

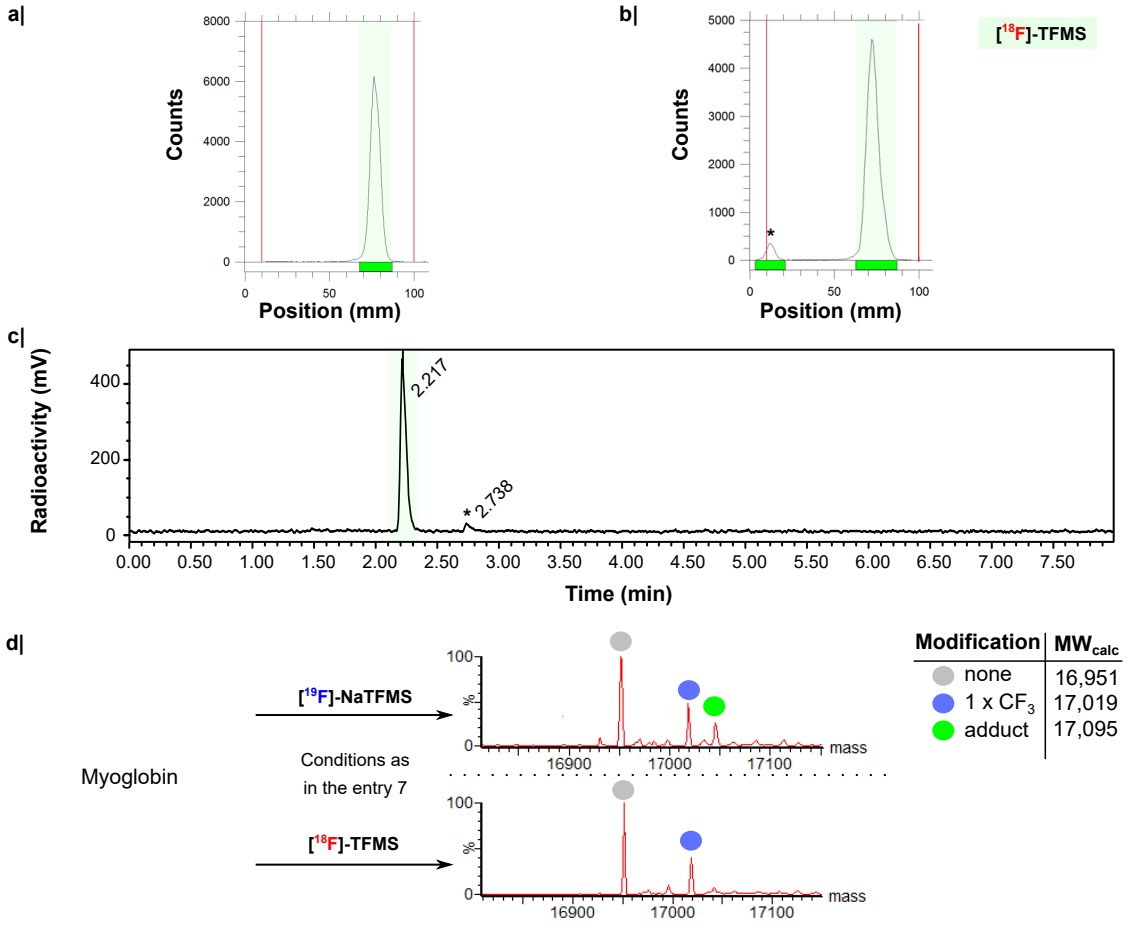


Fig. 3.10: Radiolabelling of model protein myoglobin. Radio TLCs traces of **a**| the crude **[^{18}F]-TFMS** reagent **b**| the protein reaction mixture (entry 6). **c**| Radio-chromatogram of the reaction mixture with putative traces of a radiolabelled protein marked with the asterisk. **d**| Mass spectra of the protein starting material (left) and comparison of spectra of reaction mixtures with protein labelled under similar conditions (entry 7) with 5 eq of [^{19}F]-NaTFMS (upper) or **[^{18}F]-TFMS** (bottom, analysis of the decayed reaction mixture). Both showed ~30% conversion to the trifluoromethylated protein. The peak marked in green ($\Delta m_{\text{obs}} = +95$) most likely stemmed from a sulphate adduct to the starting material protein ($\Delta m_{\text{calc}} = +96$).

Considering the difficulties in accessing high quantities [^{18}**F**]**-TFMS**, it was deemed unlikely that the protein radiolabelling protocol can be efficiently optimised in such a form and the project was concluded.

3.4 Conclusions and discussion

This chapter investigated the idea of protein radiotrifluoromethylation using the developed radical chemistry based on the trifluoromethanesulfinate reagent. Since the radiosynthesis of the ^{18}F-labelled reagent had not been reported, various synthetic routes to the radical precursors were designed and tested.

Finding a compromise between a suitable method for both alkyl sulfinate and [^{18}F]CF_3 group synthesis, a multi-step pathway was considered. Choice of the sulfinate precursor in the form of benzothiazole sulfone was dictated by highly efficient transformations leading to the sulfinate from theoretically easy to label SCF_2Br motif. This strategy was reinforced by, then, preliminary successful results achieved for ^{18}F-*difluoromethylation* of this type of substrates[321]. Moreover, the more sluggish elimination step for different (hetero)aromatic scaffolds made their choice disadvantaged[299]. For example, preliminary studies into the reactivity of the pyridyl sulfones scaffold (not shown) confirmed these notions and cemented the chosen strategy. In order to access the ^{18}F-labelled reagent ([^{18}**F**]**-NaTFMS**) for protein trifluoromethylation, the following established radiolabelling conditions were tested - i) silver-mediated halogen exchanges[298] as well as ii) a more speculative difluorocarbene mediated disulfide splitting[308]. Initiated with feasibility studies, nucleophilic fluorination of the **BzSCF$_2$Br** in the presence of silver triflate turned out to yield a product that was different

from expected (but not entirely surprising), which led to the dead end. This was most likely a consequence of the choice of specific aryl scaffold. In light of this, the focus was shifted to the alternative strategy based on a $:CF_2$ mediated approach. Although the desired **BzSCF$_3$** product was formed, the reaction performed poorly under conditions resembling those used in radiochemistry. Nevertheless, subsequent steps leading to the sulfinate showed promising results, raising hopes of translation into a semi-efficient radiosynthesis. Perhaps, these experiments should be regarded as a general successful proof-of-concept and should have directed further work into the investigations of a more productive source of the $[^{18}F]CF_3^-$ equivalent. However, considering the scope of this thesis and due to the operational and time constraints, the radiosynthesis following this strategy was pursued. Both pathways performed in a surprisingly good agreement with the feasibility studies, indicating a good prediction value of such experiments. While the silver-mediated strategy was expectedly not successful, the $:CF_2$ approach allowed construction of the **[^{18}F]-BzSCF$_3$** precursor in a limited RCY. Although further optimisations of the labelling reaction were performed, some aspects of the radiochemistry procedures were not accounted for, such as the limited $[^{18}F]F^-$ solvent transfer, leaving room for a further technical improvement - Fig. 3.13. Importantly, the total radiosynthesis was amenable for automation, highlighting a high potential of the whole sequence. Final experiments aimed at protein radiolabelling revealed a seemingly insufficient reactivity of this reagent. Use of the same reagent, but synthesised via a more elegant, single-step procedure (with $\sim$(10-20)x higher RCY)[320] for protein radiolabelling was also unsuccessful. Results from this study suggested that while radical $[^{18}F]$-trifluoromethylation of proteins is

feasible, a more reactive radical precursor than the proposed here would be essential to achieve that. It should be noted that alternative NaTFMS radical initiation methods failed (photoredox, metal catalysts - data not shown). Perhaps, a more important issue to be addressed first, is the construction of $[^{18}F]CF_3$ group. Current methods are limited in terms of RCY and molar activities, which preclude their widespread applications. Recent progress in the field[322] provides future hope that the trifluoromethyl radical could be exploited for protein radiolabelling.

3.5 Experimental

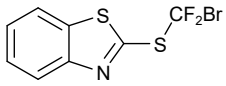
3.5.1 Feasibility studies

General information

Reagents and solvents were purchased from commercial suppliers and used without further purification. Room temperature (rt) refers to ambient temperature. For flash column chromatography Geduran^{TM} Si 60 (0.040-0.063 mm) silica gel was used. Chemical shifts are reported in parts per million (ppm) relative to the centre of the residual solvent signal. Fluorine spectra were referenced against NMR grade 2,2,2-trifluoroethanol (TFE, -77.7 ppm, t, 3F). Spectra were analysed using the Mestrelab software (MestReNova v11.0.2). Relevant 2D experiments were performed to facilitate the signal assignment. All NMR experiments for small molecules were performed on a Bruker AVIIIHD 400 nanobays (400.1 MHz for proton) running TOPSPIN 3 equipped with a 5 mm z-gradient multinuclear BBFO probe. Coupling constants (J) are reported to the nearest 0.1 Hz. Quantitative fluorine experiments were performed with TFE as the internal standard with following experimental parameters: number

of scans: 128 - 256, acquisition time: 3 s, delay time 7 s with resolution not smaller than 0.33 Hz/pt. High resolution mass spectrometry was performed with Agilent 8890 GC-TOF-MS System with the electron impact ionisation.

Synthesis of $BzSCF_2Br$



A suspension of NaH (60% dispersion in mineral oil, 0.6 g, 15 mmol) in THF (15 mL) was cooled down to 0 °C, and 2-mercaptobenzothiazole (BzSH, 1.67 g, 10 mmol) was added in one portion. After being stirred at 0 °C for 10 mins, the reaction mixture was warmed up to rt and stirred for another 30 mins. Then, it was cooled to −78 °C and CF_2Br_2 (1.1 mL, 12 mmol) was added in one portion. The reaction mixture was stirred for 18 h at rt, quenched with water (40 mL) and extracted with DCM (2 x 30 mL) and hexane (30 mL). The organic layers were combined, dried over $MgSO_4$ and evaporated in vacuo. The crude residue was dissolved in small volumen of hexane and purified via flash chromatography on silica gel (hexane/EtOAc 5%) to give **$BzSCF_2Br$** as a yellow solid (2.07 g, 70%).

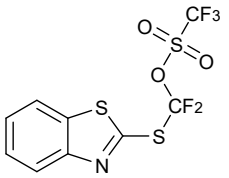
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.16 (ddd, J = 8.2, 1.4, 0.7 Hz, 1H, H4), 7.91 (ddd, J = 7.9, 1.4, 0.7 Hz, 1H, H7), 7.56 (ddd, J = 8.2, 7.3, 1.4 Hz, 1H, H5), 7.50 (ddd, J = 7.9, 7.3, 1.4 Hz, 1H, H6),

^{13}C NMR (126 MHz, $CDCl_3$) δ 154.0 (t, J = 1.8 Hz, C2), 153.2 (C3a), 138.2 (C7a), 130.9 (t, J = 322.0 Hz, CF_2Br), 127.1 (C5), 126.9 (C6), 124.5 (C4), 121.5 (C7),

^{19}F NMR (376 MHz, $CDCl_3$) δ −21.22 (s, 2F),

HRMS-(EI) $[M]^+$: Calcd: 294.8931 (100%), 296.8909 (97%), found: 294.8934 (100%) and 296.8913 (97%).

Synthesis of $BzSCF_2OTf$



AgOTf (102.8 mg, 0.4 mmol) was weighed into a glass vial to which a solution of **$BzSCF_2Br$** (59 mg, 0.2 mmol) in 1,2-dichloroethane (DCE, 0.7 mL) was added and the whole mixture was heated at 60 °C for 1 h. After that, the solvent was removed in vacuo and the remaining solid was extracted with $CDCl_3$. The instability of the compound prevented its separation from excess AgOTf. ^{19}F NMR analysis indicated over >95% consumption of the starting material. The $^5J_{F,F}$ coupling pattern has been observed for similar compounds[323].

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.30 (ddd, J = 8.2, 1.3, 0.7 Hz, 1H, H4), 7.96 (ddd, J = 7.9, 1.3, 0.7 Hz, 1H, H7), 7.64 (ddd, J = 8.2, 7.3, 1.3 Hz, 1H, H5), 7.59 (ddd, J = 7.9, 7.3, 1.3 Hz, 1H, H6),

^{13}C NMR (126 MHz, $CDCl_3$) δ 154.1 (t, J = 2.9 Hz, C2), 150.6 (C3a), 137.3 (C7a), 128.5 (C5), 128.2 (C6), 126.8 (t, J = 318.0 Hz, CF_2OTf), 123.7 (C4), 121.8 (C7), 117.8 (q, J = 322.3 Hz, CF_3),

^{19}F NMR (376 MHz, $CDCl_3$) δ −39.17 (q, J = 5.3 Hz, 2F), −72.92 (t, J = 5.3 Hz, 3F),

HRMS-(EI) $[M]^+$: Calcd: 364.9268, found: 364.9370.

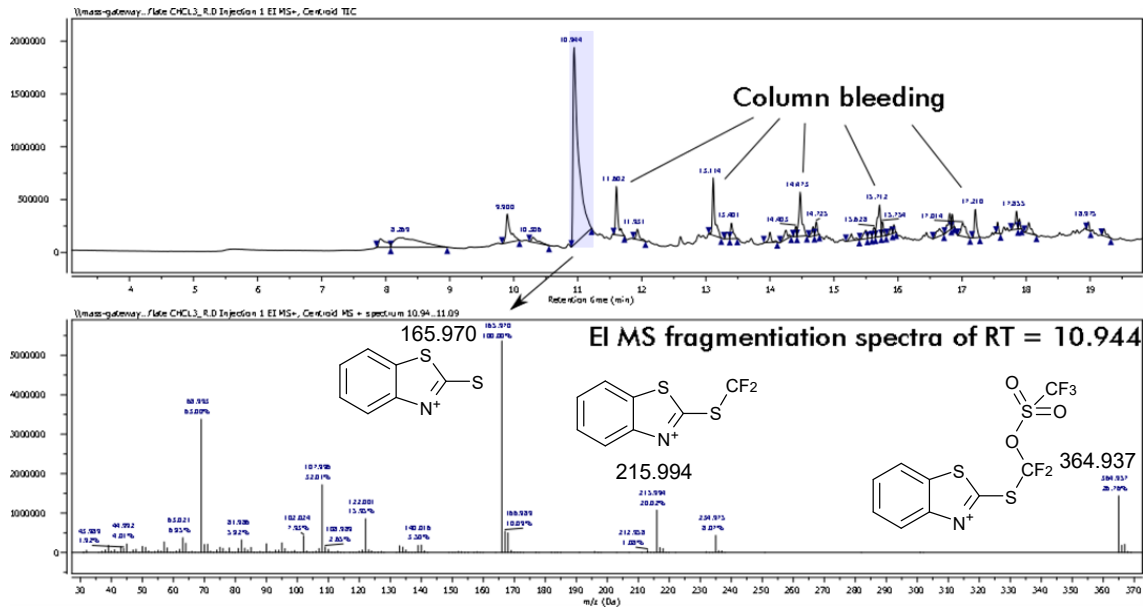


Fig. 3.11: GC-HRMS analysis of **BzSCF$_2$OTf** dissolved in $CHCl_3$. TIC chromatogram (top) and mass spectrum (bottom, combined spectra of the region highlighted in blue) supported the proposed structure upon analysis of the fragmentation pattern.

Nucleophilic fluorination of BzSCF$_2$OTf

To the crude reaction mixture containing **BzSCF$_2$OTf** (see above), KF (12.8 mg, 0.22 mmol) and a phase transfer catalyst (dicyclohexano-18-crown-6, 82 mg, 0.22 mmol) were added and the mixture was heated for 30 mins at 60 °C. After that, an aliquot (100 μL) was diluted in $CDCl_3$ for ^{19}F NMR analysis.

Nucleophilic fluorination of BzSCF$_2$Br

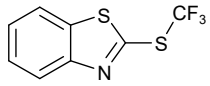
To an oven dried round bottom flask, KF (12.8 mg, 0.22 mmol) and a phase transfer catalyst (dicyclohexano-18-crown-6, 82 mg, 0.22 mmol) were added and the solids were dissolved in anhydrous MeCN (1 mL). The mixture was heated up to 60 °C and the solvent was evaporated under a stream of N_2. After that, the precursor **BzSCF$_2$Br** (64 mg, 0.22 mmol) and the solvent (1 mL of anhydrous MeCN, DMF

or DCE) were added. The mixture was heated for 30 mins at a desired temperature (70 °C for MeCN and DCE or 120 °C for DMF). After cooling, an aliquot (100 μL) was diluted in $CDCl_3$ for the ^{19}F NMR analysis.

Synthesis of $BzSCF_3$ and $BzSNCF_2H$

KF (250 mg, 4.3 mmol), **BzSSBz** (711 mg, 2.15 mmol) and CuI (513 mg, 2.7 mmol) were suspended in DMF (20 mL). Then, methyl chlorodifluoroacetate was added (F_2ClCCO_2Me, 0.45 mL, 4.3 mmol) under nitrogen atmosphere. The reaction mixture was heated to 135 °C for 30 mins and then, was allowed to cool to rt. The crude mixture was partitioned between DCM and brine. The aqueous phase was extracted with DCM, the organic phases combined, dried over $MgSO_4$, concentrated and purified by flash chromatography on silica gel (hexane:DCM 1:1 to 1:3). This allowed isolation of **$BzSCF_3$** (40 mg, 8%) as a colourless oil as well as **$BzSNCF_2H$** (140 mg, 30%) as a yellow oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.11 (ddd, J = 8.2, 1.4, 0.7 Hz, 1H, H4), 7.88 (ddd, J = 8.0, 1.4, 0.7 Hz, 1H, H7), 7.54 (ddd, J = 8.2, 7.2, 1.4 Hz, 1H, H5), 7.47 (ddd, J = 8.0, 7.2, 1.4 Hz, 1H, H6),



$BzSCF_3$

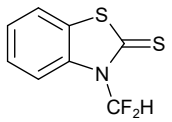
^{19}F NMR (376 MHz, $CDCl_3$) δ −40.55 (s, 3F),

^{13}C NMR (126 MHz, $CDCl_3$) δ 153.1 (C3a), 151.7 (q, J = 3.0 Hz, C2), 137.9 (C7a), 128.2 (q, J = 311 Hz, CF_3), 127.0 (C5), 126.7 (C6), 124.1 (C4), 121.3 (C7),

HRMS (EI) $[M]^+$: Calcd: 234.9732, found: 234.9901.

The characterisation conforms to data from literature[177].

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.16 (t, J = 57.9 Hz, 1H, CF_2H) 7.71 – 7.64 (bd, J = 8.5 Hz, 1H, H4), 7.47 (dd, J = 8.0, 1.4, Hz, 1H, H7), 7.42 (ddd, J = 8.5, 7.6, 1.4 Hz, 1H, H5), 7.36 (td, J = 8.0, 7.6, 1.4 Hz, 1H, H6),



BzSNCF$_2$H

^{19}F NMR (376 MHz, $CDCl_3$) δ −106.23 (dd, J = 57.6, 1.9 Hz, 2F)

^{13}C NMR (126 MHz, $CDCl_3$) δ 191.5 (t, J = 3.0 Hz, C2), 137.3 (t, J = 1.8 Hz, C3a), 127.5 (C5), 126.1 (C7a), 125.9 (C6), 121.4 (C4), 113.8 (t, J = 4.5 Hz, C7), 110.0 (t, J = 253.6 Hz, CF_2H),

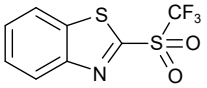
HRMS (EI) $[M]^+$: Calcd: 216.9826 , found: 216.9828.

The characterisation conforms to data from literature[312].

General protocol for BzSSBz trifluoromethylation optimisation

A glass vial was charged with **BzSSBz** (50 mg, 0.15 mmol), alkali fluoride salt (0.3 mmol), Cu(I) catalyst (0.03 mmol), any other additive (0.03 mmol) and a stirring bar. After sealing with a septum, the vial was flushed with nitrogen and anhydrous DMF (0.7 mL) was added. The difluorocarbene precursor (0.3 mmol) was added to this suspension, and the reaction mixture was heated to a desired temperature. After 20 min, the vial was cooled to rt and the suspension was transferred to a microcentrifuge tube, together with the vial rinse (0.7 mL of $CDCl_3$). The organic phases were combined and centrifuged (15,000 rpm, 2 mins). The supernatant (549 μL) was analysed by quantitative ^{19}F NMR after addition of the internal standard (1.1 μL of TFE). Concentration of species at −40.55 ppm (**BzSCF$_3$**) and −106.23 ppm (**BzNCF$_2$H**) was determined by intensity comparison with the standard.

Synthesis of $BzSOOCF_3$



$BzSCF_3$ (112 mg, 0.48 mmol) was dissolved in MeCN (10 mL) and the solution was placed on ice. Then, an aqueous $NaIO_4$ solution (1020 mg, 4.8 mmol, 10 eq, 10 mL H_2O) and $RuCl_3$ (15 mg, <15 mol%) were added. The cooling bath was removed after 20 mins, and the reaction mixture was stirred at rt for additional 3 hours. The crude reaction mixture was partitioned between DCM and water. The aqueous layer was extracted 3x with DCM, the organic layers were combined, dried over $MgSO_4$, concentrated and purified via flash chromatography on silica gel (hexane:DCM up to 40%) to yield a white solid (92 mg, 72%).

^{1}**H NMR** (400 MHz, $CDCl_3$) δ 8.38 - 8.36 (m, 1H, H4), 8.12 - 8.09 (m, 1H, H7), 7.76 – 7.69 (m, 2H, H5&H6),

^{19}**F NMR** (376 MHz, $CDCl_3$) δ −76.04 (s, 3F),

^{13}**C NMR** (126 MHz, $CDCl_3$) δ 157.1 (q, J = 2.6 Hz, C2), 153.2 (C3a), 138.5 (C7a), 129.7 (C5), 128.7 (C6), 126.8 (C4), 122.5 (C7), 119.7 (q, J = 327.1 Hz, CF_3),

HRMS (EI) $[M]^+$: Calcd: 266.9636, found: 266.9656.

The characterisation conforms to data from literature[177].

Screening conditions for crude $BzSCF_3$ oxidation

A glass vial was charged with **BzSSBz** (50 mg, 0.15 mmol), KF (17.5 mg, 0.3 mmol), CuI (5.7 mg, 0.03 mmol), TMEDA (4.5 μL, 0.03 mmol) and a stirring bar. After sealing with septum, the vial was flushed with nitrogen and anhydrous DMF (0.7 mL) was added. The difluorocarbene precursor (methyl chlorodifluoroacetate - 31.6

μL, 0.3 mmol) was added to this suspension, and the reaction mixture was heated to 130 °C. After 20 mins, the vial was cooled to rt and part of the crude mixture was withdrawn with a syringe (350 μL) to a microcentrifuge tube open to air. To this solution aqueous $NaIO_4$ was added (e.g. 80 mg/mL, 350 μL for ∼7 eq), solid $RuCl_3$ (e.g. 5 mg for ∼0.8 eq) and an additional solvent (e.g. 300 μL MeCN). The mixture was incubated for 20 mins after which half of the reaction mixture was withdrawn. To that solution MeCN-d3 (100 μL) as well as the internal standard (1.1 μL of TFE) was added and the mixture was analysed by quantitative ^{19}F NMR.

C18 cartridge trapping - feasibility studies

Trifluoromethylation and oxidation were performed as described above, on a 0.15 mmol scale with respect to **BzSSBz** and with sufficient excess $RuCl_3$ and $NaIO_4$ i.e. above 2 eq and 30 eq respectively. The crude reaction mixture (∼2 mL of aq. DMF solution) was centrifuged (15,000 xg, 2 mins) and the supernatant was diluted with water (8 mL). The C18 cartridge (Sep-Pak C18 Plus Short Cartridge, 360 mg Sorbent) was pre-conditioned with EtOH (2 mL) and H_2O (4 mL). The suspension of **BzSOOCF$_3$** was slowly passed through the cartridge during 5 mins. The cartridge was then washed with 10% aq. EtOH (2 mL) before eluting **BzSOOCF$_3$** with EtOH (2 mL). ^{19}F NMR analysis after a further elution with an additional portion of EtOH (2 mL) confirmed that the desired compound was completely removed from the cartridge.

Trifluoromethanesulfinate elimination from $BzSOOCF_3$

To an EtOH solution of **$BzSOOCF_3$** after the C18 cartridge purification, solid $NaBH_4$ (14.4 mg, 0.38 mmol, > 10 eq) was added. The mixture was incubated for 5 mins at rt after which an aliquot (supplemented with 50 μL of D_2O) was analysed by ^{19}F NMR. This indicated the complete consumption of the starting material and a clean conversion to trifluoromethanesulfinate as confirmed by spiking with the authentic reference.

^{19}F NMR (376 MHz, D_2O) δ −88.5 (s, 3F).

3.5.2 Radiochemistry

General information

Work involving radiochemistry was conducted at the Centre for Cyclotron Research, University of Liege, Belgium. All manual radiochemical reactions were carried out in sealed glass vials. All automated syntheses were performed in a dedicated reaction vessels (FASTLab).

Analytical LC-Radio-UV was performed with Waters Acquity UPLC H-Class system equipped with Waters PDA detector, connected in series to a NaI(Tl) scintillation detector (Eberline) for the radioactivity measurement. The whole analytical suite was controlled by Empower 2 software. LC separation was achieved with Waters BEH C18 column (1.7 μm, 50 x 2.1 mm) with a flow rate of 0.5 mL/min starting with an isocratic elution at 20% aq MeCN for (0 → 0.5) min, followed by a gradient up to 100% MeCN (0.5 → 6) min, a wash step with 100% MeCN for (6 → 6.95) mins and finally, the system was reconditioned with 20% MeCN (6.95 → 8) min. Both aque-

ous and MeCN mobile phases contained 0.1% formic acid. For [^{18}**F**]**-TFMS** reagent analysis, an additive in the form of an ion pairing reagent was added to the mobile phases (replacing formic acid): 5 mM tetrabutylammonium bromide (NBu_4Br).

The reaction mixtures were additionally analysed with radioTLC (Polygram SIL G/UV254, AR200 Berthold TLC scanner) eluted either with $CHCl_3$ or MeOH. The radiochemical purity was confirmed by an UPLC analysis, while the purity of the final compounds was verified by the measurement of UV absorbance spectrum (220 nm). Radiochemical conversions are reported non-decay corrected, unless stated otherwise. Radioactivity of the samples for the molar activity determination was measured in a dose calibrator (Veenstrat).

Semi-preparative RP-HPLC was realised with Waters 600 pump system (with He degassing) coupled to a Waters 996 PDA detector, using a XTerra C18 column (7 μm, 300 x 7.8 mm) and a 8-mL injection loop (Rheodyne injector). Purification of the reagent was realised at a flow rate of 5 mL/min and an isocratic elution with 20% MeCN (5 mM NBu_4Br in water). The elution profile was also monitored with a custom-made Geiger-Müller radioactivity counter.

All of the utilised cartridges - Sep-Pak Light C18 (360 mg, 55-105 μm) and Accell Plus QMA Carbonate Plus Light (46 mg, 37-55 μm) were obtained from Waters. The C18 solid phase extraction cartridge was activated beforehand with MeCN (2 mL) and then reconditioned with water (6 mL) before use for trapping.

[^{18}F]F^{-} preparation

Non carrier added ^{18}F-fluoride was produced by 18 MeV proton beam irradiation (varied time IBA cyclone 18/18 cyclotron) of >95% enriched ^{18}O water (Rotem or ABX) leading to ^{18}O(p,n)^{18}F nuclear reaction in a niobium target. At the end of the bombardment, the target contents were delivered through ∼75 m small-bore Teflon tubing (1.6 mm x 0.8 mm) directly to the FASTlab synthesizer in the hot cell, where the ^{18}F-fluoride was trapped an ion exchange cartridge (QMA Carbonate Cartridge) and the enriched water was recovered in a separate vial.

The elution of radioactivity from the cartridge into the reaction vessel was achieved with a mixture of Kryptofix 2.2.2 (4,7,13,16,21,24-hexaoxa-1,10-diazabicyclo[8.8.8] hexacosane - 7.5 mg in 0.6 mL MeCN) and K_2CO_3 (1.4 mg in 0.15 mL of water). The mixture was azeotropically removed by heating up to 120 °C under a flow of nitrogen and then dried under vacuum. The dried ^{18}F[KF]-cryptand complex was then either used in a labelling reaction upon the addition of reagents, or re-dissolved in (0.5 - 2) mL of a desired solvent and transferred via an output line into a glass vial outside the hot cell for a further manipulation during manual screenings.

Manual radiosynthesis - screening

For manual screening the solid reagents (and additives) were weighed into a glass vial: **BzSSBz** (7 mg, 21 μmol) and CuBr (11 mg, 80 μmol), after which they were dissolved in DMF solution (0.7 mL) of ClF_2CCO_2Me (6 μL, 57 μmol) and TMEDA (12 μL, 80 μmol). The vial was then sealed and an aliquot of ^{18}F-fluoride in DMF (∼(50 - 100) μL) was added using a syringe, followed by the measurement of the

starting radioactivity. The mixture was then heated on a metal heating block for required time. Temperature was controlled by a thermocouple home made system. After that, the vial was cooled to rt, unsealed and the contents were analysed by radioTLC and UPLC. To investigate the oxidation step, the reaction mixture was diluted with water (8 mL) and the suspension was slowly passed through a pre-activated C18 cartridge. The adsorbed compounds were then eluted with MeCN (1 mL) and solution of oxidants ($RuCl_3$, 0.6 mg, 3 μmol and $NaIO_4$, 19 mg, 90 μmol in 0.35 mL of water) was added. The reaction mixture was analysed after 5 mins using the previously mentioned methods. To test the elimination step, the sulfone product was again trapped and eluted from a C18 cartridge in an analogous way. $NaBH_4$ (5 mg, 132 μmol in 0.5 mL EtOH) was added to the eluate and the reaction mixture was analysed in a standard way after 5 mins.

Automated radiosynthesis

The automated radiosynthesis was performed on a FASTlab synthesizer using standard cassette (fluorodeoxyglucose phosphate cassette template). The solutions of reagents and solvents were prepared in small sealed glass vials which were positioned on the FASTlab manifold. Typical casette scheme is shown in Fig. 3.12 and the full, optimised synthesis sequence can be found in the Appendix (Fig. 3.14).

The sequence began with preparation of ^{18}F-fluoride and its delivery to the reaction vessel as described before. Then, without cooling the vessel, the labelling reagents: **BzSSBz** (3.5 mg, 10 μmol), CuBr (3 mg, 21 μmol), TMEDA (3 μL, 20 μmol) and ClF_2CCO_2Me (4.2 μL, 40 μmol) in DMF (1.6 mL) were added through

the central tubing and the radiolabelling reaction was then performed for 10 mins. After that, the reaction mixture was diluted three times in the syringe with H_2O (∼8 mL in total) and the product was trapped on a C18 cartridge. The reaction vessel was additionally washed with a portiof of H_2O (∼4 mL) that was also passed through the cartridge. The oxidation step was performed by slow application of the reagents solution (2.4 mg, 12 μmol $RuCl_3$ and 80 mg, 375 μmol $NaIO_4$ in 4 mL of 5% aqueous MeCN) to the cartridge, with a 60 s pause after every 1 mL flowed, leading to the total reaction time of ∼5 mins. The crude sulfone product was eluted from the C18 cartridge with MeCN (1 mL) into an outlet vial, to which the $NaBH_4$ solution was added (5 mg, 132.5 μmol in 0.5 mL of EtOH). The crude reaction mixture with the desired product was then diluted with water (8 mL) and purified via semi-preparative HPLC. The fraction with the ^{18}F-trifluoromethanesulfinate was then collected into a glass vial (RT = (7 - 8) mins) and directly applied to an pre-activated C18 cartridge. The final pure compound was recovered by elution with MeCN (0.5 mL). This solution was concentrated by evaporation under a flow of nitrogen ((5 - 15) mins) and diluted with a desired quantity of the buffer (0.1 M NH_4OAc pH 6) for further use in protein reactions.

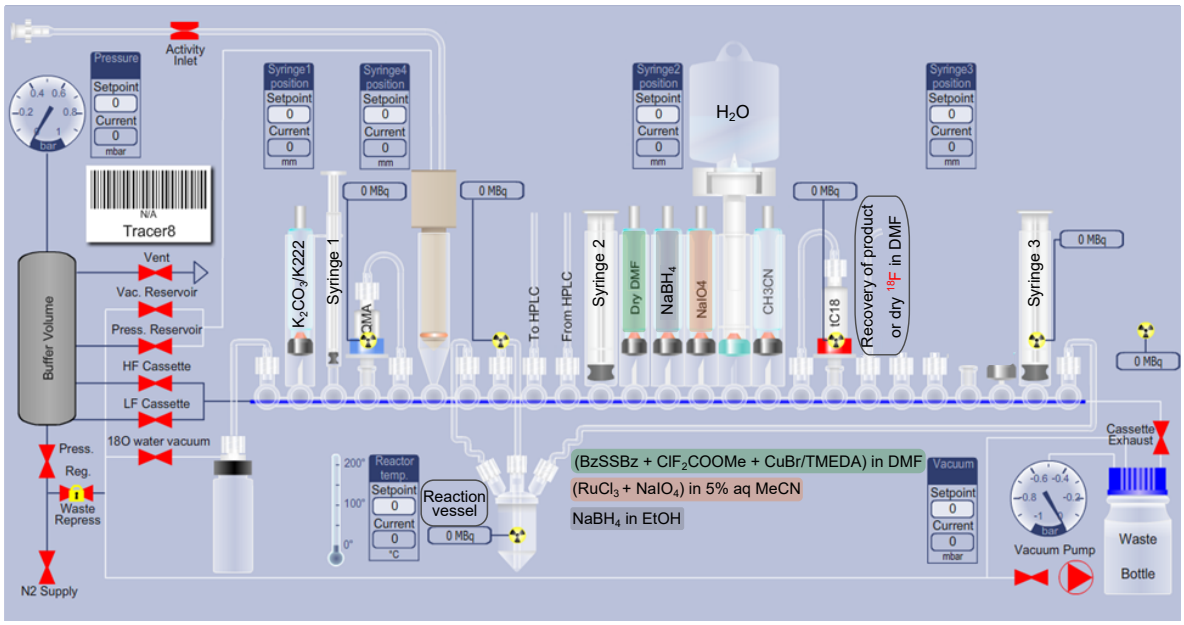


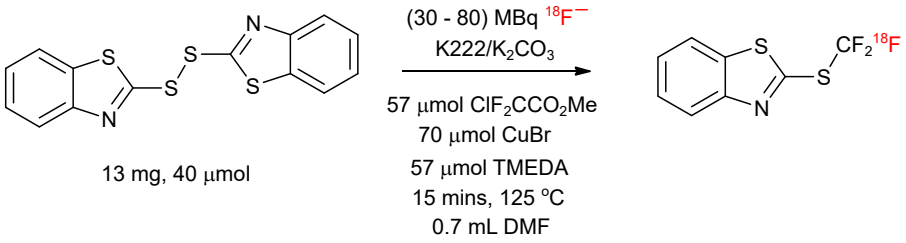
Fig. 3.12: Cassete schematic with positions of cartridges and reagents. Solutions of relevant reagents are highlighted: radiolabelling reagents - green, oxidation reagents - orange and the elimination reagent - grey. Additional vials contained MeCN and K222/K_2CO_3 solutions.

Protein radiolabelling

To a solution of eqMb (200 μL, 10 mg/mL in aq. 0.1 M NH_4OAc buffer pH 6) in a glass vial, an aliquot of aqueous [^{18}**F**]-**TFMS** (<50 μL, (3.5 - 20) MBq, prepared as above) was added and the reaction was initiated by addition of *t*BuOOH (720 mM in 0.1 M NH_4OAc buffer pH 6). The reaction mixture was heated if needed and after a desired time assayed with radio-TLC and UPLC.

3.6 Appendix

Solvent and additives screening - :CF_2 radiotrifluoromethylation



Tab. 3.11: **BzSSBz** radiolabelling in different solvents and in the presence of various additives. DMA = dimethylacetamide, NMP = *N*-methyl-2-pyrrolidone, PC = propylene carbonate, NMM = *N*-Methylmorpholine, TED = triethylenediamine, TMPDA = *N,N,N',N'*-tetramethyl-1,3-propanediamine

Entry	Solvent	Additive	TLC	UPLC	RCC
1	DMF	–	17%	37%	6%
2[a]	MeCN	–	7%	21%	1%
3	DMA	–	7%	–	–
4	NMP	–	3%	–	–
5	PC	–	5%	–	–
6	NMM	–	12%	–	–
7	DMF	TED[b]	55%	–	–
8	DMF	TMPDA[b]	10%	–	–
9	DMF	AcOK	6%	–	–
10	DMF	AgOTf	7%	–	–
11	DMF	PPh_3	28%	–	–
12	DMF	*t*BuOK	4%	–	–
13	DMF	NBu_4Br	25%	20%	5%
14	DMF	K_2CO_3	32%	19%	6%

[a] - temperature = 90 °C, [b] - additive in place of TMEDA

Radioactivity to solvent transfer

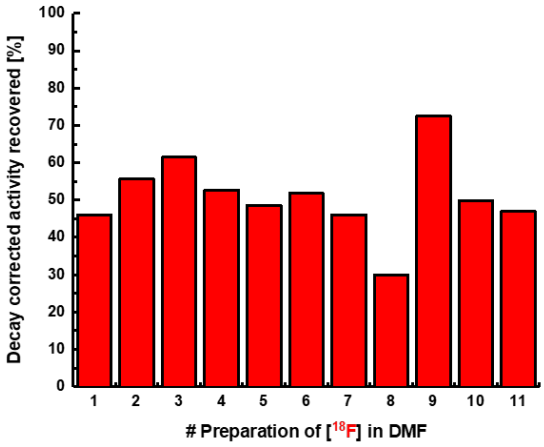


Fig. 3.13: Decay corrected radioactivity recovery in the chosen solvent - DMF. $[^{18}F]F^-$ was prepared in the synthesiser according to the standard procedure. After evaporation to dryness, it was redissolved in 1.5 mL of DMF. The quantified activity refers to that of $[^{18}F]F^-$ DMF solution as compared with the starting activity recovered from the QMA cartridge. It was found that only $\sim$50% of the expected value was obtained. The residual activity was found within the reaction vessel.

Automation program details

Step	Message	Duration	Syring...	Syring...	Syringe 3	Pressu...	Vacuum	Reactor temp.
1	Starting point	1	0	0	0	800	-100	0
2	Flushing the manifold	5	0	0	0	800	-100	0
3	Flushing the manifold	1	0	0	0	800	-100	0
4	Flushing the manifold	1	0	0	600	800	-100	0
5	Flushing the manifold	1	0	0	100	800	-100	0
6	Flushing the manifold	5	0	0	100	800	-100	0
7	Flushing the manifold	1	0	0	100	800	-100	0
8	Flushing the manifold	1	0	0	0	800	-100	0
9	Flushing the manifold	1	0	0	0	800	-100	0
10	Flushing the manifold	1	0	0	600	800	-100	0
11	Flushing the manifold	1	0	0	100	800	-100	0
12	Flushing the manifold	5	0	0	100	800	-100	0
13	Flushing the manifold	1	0	0	100	800	-100	0
14	Flushing the manifold	1	0	0	0	800	-100	0
15	Flushing the manifold	25	0	0	0	800	-100	0
16	ACN vial pressurisation	5	0	0	0	800	-100	0
17	ACN vial pressurisation	15	0	0	0	800	-100	0
18	Acetonitrile vial pressurisation	5	0	0	0	800	-100	0
19	NaIO4 vial pressurisation	5	0	0	0	800	-100	0
20	tC18 Cartridge conditioning	5	0	0	0	800	0	0
21	tC18 cartridge conditioning	4	0	0	0	800	-200	0
22	tC18 cartridge conditioning	1	0	0	0	800	0	0
23	Flushing the manifold	10	0	0	0	800	-200	0
24	Acetonitrile vial repressurisation	10	0	0	0	800	-200	0
25	Acetonitrile vial repressurisation	5	0	0	0	800	-200	0
26	Ready to receive activity	1	0	0	0	800	-200	0
27	Activity trapping on QMA	2	0	0	0	500	-400	0
28	Activity trapping on QMA	25	0	0	0	500	-400	0
Step	Message	Duration	Syring...	Syring...	Syringe 3	Pressu...	Vacuum	Reactor temp.
29	Activity trapping on QMA	1	0	0	0	500	-400	0
30	Activity trapping on QMA	25	0	0	0	500	-400	0
31	Activity trapping on QMA	1	0	0	0	500	-400	0
32	Activity trapping on QMA	1	0	0	0	500	-400	0
33	Activity trapping on QMA	30	0	0	0	500	-400	0
34	Activity trapping on QMA	1	0	0	0	500	-400	0
35	Activity trapping on QMA	25	0	0	0	500	-400	0
36	Activity trapping on QMA	1	0	0	0	500	-400	0
37	Activity trapping on QMA	10	200	0	0	600	-500	0
38	Activity trapping on QMA	20	200	0	0	700	-500	0
39	Activity trapping on QMA	30	200	0	0	1000	-500	0
40	Activity trapped on QMA	5	200	0	0	700	-200	0
41	Eluent transferred to syringe 1	2	200	0	0	0	-200	0
42	Eluent transferred to syringe 1	1	600	0	0	0	-200	0
43	Elution of activity into reactor	5	600	0	0	0	-200	0
44	Elution of activity into reactor	5	500	0	0	0	-200	0
45	Elution of activity into reactor	5	400	0	0	0	-200	0
46	Elution of activity into reactor	5	300	0	0	0	-200	90
47	Elution of activity into reactor	5	200	0	0	0	-200	90
48	Elution of activity into reactor	5	0	0	0	0	-200	90
49	Elution of activity into reactor	15	0	0	0	20	-200	105
50	Evaporation	300	0	0	0	130	-1000	120
51	Evaporation	5	0	0	0	200	-1000	120
52	Evaporation	30	0	0	0	200	-1000	120
53	Evaporation	10	0	0	0	200	-100	105
54	Evaporation	10	0	0	0	50	-150	105
55	Evaporation	7	0	0	0	80	-1000	120
56	Evaporation	1	0	300	0	80	-1000	120

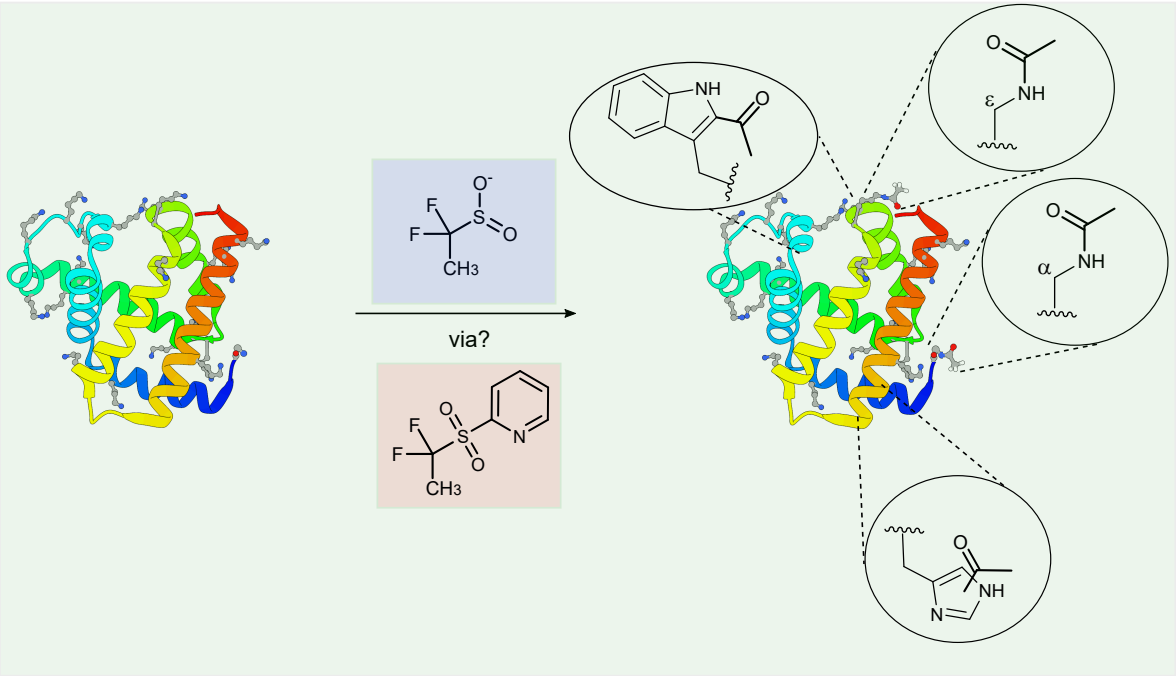
Step	Message	Duration	Syring...	Syring...	Syringe 3	Pressu...	Vacuum	Reactor temp.
57	Evaporation	120	0	300	0	80	-1000	120
58	Evaporation	15	0	300	0	200	-100	125
59	Drying ACN transfer	2	0	300	0	500	-100	125
60	Drying ACN transfer	2	0	150	0	300	-100	125
61	Drying ACN transfer	3	0	0	0	50	-100	125
62	Evaporation	7	0	0	0	200	-1000	125
63	Labelling + flushing the manifold	10	0	0	0	700	-300	125
64	Labelling	1	0	0	0	700	-700	125
65	Labelling + tC18 cart conditioning	7	0	0	0	700	-700	125
66	Labelling + flushing the manifold	20	0	0	0	700	-300	125
67	Labelling	600	0	0	0	300	-300	125
68	Labelling	60	0	0	0	300	-300	85
69	Labelling	10	0	240	0	300	-50	85
70	Dilution prior to tC18 trapping	1	0	240	0	300	-50	0
71	Dilution prior to tC18 trapping	1	0	115	0	300	-50	0
72	1	1	0	0	0	300	-50	0
73	Dilution prior to tC18 trapping	2	0	600	0	300	-50	0
74	Dilution prior to tC18 trapping	1	0	0	0	300	-50	0
75	Dilution prior to tC18 trapping	1	0	0	0	300	-50	0
76	Dilution prior to tC18 trapping	5	0	400	0	300	-50	0
77	Dilution prior to tC18 trapping	1	0	400	0	300	-300	0
78	Dilution prior to tC18 trapping	7	0	600	0	300	-300	0
79	Trapping on tC18 cartridge (1st batch)	4	0	600	0	300	-300	0
80	Trapping on tC18 cartridge (1st batch)	4	0	450	0	300	-300	0
81	Trapping on tC18 cartridge (1st batch)	10	0	300	0	300	-300	0
82	Trapping on tC18 cartridge (1st batch)	5	0	150	0	300	-300	0
83	Trapping on tC18 cartridge (1st batch)	4	0	0	0	300	-300	0
84	Dilution prior to tC18 trapping	1	0	0	0	300	-300	0
Step	Message	Duration	Syring...	Syring...	Syringe 3	Pressu...	Vacuum	Reactor temp.
85	Dilution prior to tC18 trapping	5	0	500	0	300	-300	0
86	Dilution prior to tC18 trapping	1	0	500	0	300	-300	0
87	Dilution prior to tC18 trapping	1	0	0	0	300	-300	0
88	Dilution prior to tC18 trapping	7	0	600	0	300	-300	0
89	Trapping on tC18 cartridge (2nd batch)	1	0	600	0	300	-300	0
90	Trapping on tC18 cartridge (2nd batch)	1	0	450	0	300	-300	0
91	Trapping on tC18 cartridge (2nd batch)	10	0	300	0	300	-300	0
92	Trapping on tC18 cartridge (2nd batch)	5	0	150	0	300	-300	0
93	Trapping on tC18 cartridge (2nd batch)	1	0	0	0	300	-300	0
94	Rinsing the reactor with water	1	0	0	0	300	-500	0
95	Rinsing the reactor with water	1	0	400	0	300	-500	0
96	Rinsing the reactor with water	1	0	400	0	300	-500	0
97	Rinsing the reactor with water	1	0	0	0	300	-500	0
98	Rinsing the reactor with water	7	0	600	0	300	-500	0
99	Trapping on tC18 cartridge (3rd batch)	1	0	600	0	300	-500	0
100	Trapping on tC18 cartridge (3rd batch)	1	0	450	0	300	-500	0
101	Trapping on tC18 cartridge (3rd batch)	10	0	300	0	300	-500	0
102	Trapping on tC18 cartridge (3rd batch)	5	0	150	0	300	-500	0
103	Trapping on tC18 cartridge (3rd batch)	1	0	0	0	300	-500	0
104	Pause waiting if you want oxidation	30	0	0	0	300	-500	0
105	Oxidation on the tC18 cartridge	1	0	0	0	300	-500	0
106	Oxidation on the tC18 cartridge	5	0	600	0	300	-500	0
107	Oxidation on the tC18 cartridge (1)	1	0	600	0	300	-500	0
108	Oxidation on the tC18 cartridge (1)	1	0	550	0	300	-500	0
109	Oxidation on the tC18 cartridge (2)	5	0	500	0	300	-500	0
110	Oxidation on the tC18 cartridge (2)	60	0	450	0	300	-500	0
111	Oxidation on the tC18 cartridge (2)	5	0	400	0	300	-500	0
112	Oxidation on the tC18 cartridge (2)	60	0	350	0	300	-500	0

Step	Message	Duration	Syring...	Syring...	Syringe 3	Pressu...	Vacuum	Reactor temp.
113	Oxidation on the tC18 cartridge (2)	5	0	300	0	300	-500	0
114	Oxidation on the tC18 cartridge (2)	60	0	250	0	300	-500	0
115	Oxidation on the tC18 cartridge (2)	5	0	200	0	300	-500	0
116	Oxidation on the tC18 cartridge (2)	60	0	200	0	300	-500	0
117	Oxidation on the tC18 cartridge (2)	5	0	200	0	300	-500	0
118	Oxidation on the tC18 cartridge (2)	5	0	150	0	300	-500	0
119	Oxidation on the tC18 cartridge (2)	5	0	100	0	300	-500	0
120	Oxidation on the tC18 cartridge (2)	5	0	50	0	300	-500	0
121	Oxidation on the tC18 cartridge (2)	5	0	0	0	300	-500	0
122	Washing of the syringe	1	0	0	0	300	-500	0
123	Washing of the syringe	1	0	300	0	300	-500	0
124	Washing of the syringe	1	0	300	0	300	-500	0
125	Washing of the syringe	4	0	0	0	300	-500	0
126	Washing of the syringe	1	0	0	0	300	-500	0
127	Washing of the syringe	1	0	200	0	300	-500	0
128	Washing of the syringe	1	0	200	0	300	-500	0
129	Washing of the syringe	3	0	300	0	300	-500	0
130	Washing of the syringe	1	0	300	0	300	-500	0
131	Washing of the syringe	10	0	0	0	300	-500	0
132	Washing of the syringe	1	0	0	0	300	-500	0
133	Washing of the syringe	1	0	250	0	300	-500	0
134	Washing of the syringe	1	0	500	0	300	-500	0
135	Washing of the syringe	1	0	500	0	300	-500	0
136	Washing of the syringe	1	0	250	0	300	-500	0
137	Washing of the syringe	4	0	0	0	300	-500	0
138	Drying of tC18 cartridge	30	0	0	0	300	-500	0
139	Drying system	30	0	0	0	300	-500	0
140	Acetonitrile elution	1	0	0	0	300	0	0
Step	Message	Duration	Syring...	Syring...	Syringe 3	Pressu...	Vacuum	Reactor temp.
141	Acetonitrile elution	5	0	100	0	300	0	0
142	Acetonitrile elution	5	0	200	0	300	0	0
143	Acetonitrile elution	1	0	200	0	300	0	0
144	Acetonitrile elution	5	0	100	0	300	0	0
145	Acetonitrile elution	5	0	50	0	300	0	0
146	Acetonitrile elution	5	0	0	0	300	0	0
147	Dillution with water	13	0	0	0	400	0	0
148	Line flush	10	0	0	0	400	0	0
149	Pause	1200	0	0	0	500	0	0

Fig. 3.14: Excerpts from the automation program highlighting critical parameters of the radiosynthesis. The columns (from left) indicate: step number, description, duration of each step [seconds], position of syringes (1-3) [mm], applied pressure of nitrogen stream [mbar], applied vacuum [mbar] and reaction vessel temperature [°C]. Note that the valves positions are not shown.

Chapter 4

Protein modifications with difluoroethyl radical precursors



Abstract

Bioconjugation methods that allow for the formation of specific protein constructs are invaluable for chemical biology experiments. Chapter 4 tests a strategy based on fluoroalkyl radical chemistry, with difluoroethylation as the model transformation. The initial approach using the sulfinate/oxidant system turned out to yield unexpected products, which were thoroughly characterised and formation of which was linked to the hydrolytic instability of the difluoroethylsulfinate precursor under oxidative conditions. It was proposed that the formation of acetyl fluoride was the most plausible explanation for the observed non-specific protein acetylation. The discovery was briefly explored as a mild protein acylation protocol. However, to overcome this side reactivity problem, an alternative platform for radical generation was established. A photoredox catalysed difluoroethylation, based on the pyridyl sulfone radical precursor, was extensively studied for biomolecular substrates. This allowed the establishment of the lower reactivity of the CF_2CH_3 radical, as well as the hydrolytic instability of the difluoroethylated tryptophan, yielding spontaneously the corresponding ketone.

4.1 Introduction

The next part of the project was concerned with the investigation of different fluoroalkyl radicals reactivity, seeking possible applications in a simultaneous bioconjugation and fluorine tagging. Having observed facile modifications with the trifluoromethyl radical, coupled with an appealing selectivity pattern, it was wondered whether a similar transformation can take place using a radical where a fluorine atom has been substituted with an R group, preferably an alkyl payload. Such a strategy would add to the increasing number of chemistries aimed at heteroaromatics modification - Chapter 1.1.2, potentially allowing access to mimics of natural PTMs and dual structure/function probes. In the view of previous results (Chapter 2), the key questions that emerged involved (i) AA residue reactivity, (ii) chemoselectivity and (iii) the influence of steric bulk versus aromatic residues solvent accessibility.

At the start of the project, there were no comprehensive reports on the reactivity of fluoroalkyl radicals towards peptides and/or proteins. Therefore, to establish a suitable platform, the same system of sulfinates as radical precursors was chosen. As such, the studies began with an investigation of the commercially available model compound - sodium 1,1-difluoroethanesulfinate (NaDFES). This chapter first presents the studies of the observed N-reactivity of NaDFES with peptide and protein substrates, together with the mechanistic proposal. Then, it examines C-difluoroethylation on a small molecule model, where hydrolytic instability of the products is investigated. Moving on, a new system for $\bullet CF_2Me$ generation is introduced as well as its application to a model protein. Finally, protein difluoroethylation is critically discussed.

4.2 NaDFES as an N-acetylating agent

4.2.1 Initial observations

In the initial experiment, NaDFES was reacted with a model protein myoglobin (eqMb) under conditions developed for sulfinate mediated trifluoromethylation[232]. Surprisingly, instead of the expected mass shift of $\Delta m = +64$ (observed only in a very low level $< 5\%$), LC-MS analysis revealed a pattern of sequential $\Delta m = +42$ additions - Fig. 4.1a. Interestingly, the modification was radical in its nature, as it was completely inhibited by the presence of a persistent radical, TEMPOL (data not shown).

In order to elucidate the sites of modification, the protein sample was digested with trypsin, and the resultant peptides were analysed via LC-MS/MS - Fig. 4.1b. The spectra of the fragmented peptides were analysed for a $\Delta m = +42$ modification on every amino acid, which revealed a preferential modification of several residues: N-terminus, Lys, Trp and His. Two possible reactive pathways were then hypothesised, involving residues from different classes: one with nucleophilic free amines, and the other with heteroaromatic rings.

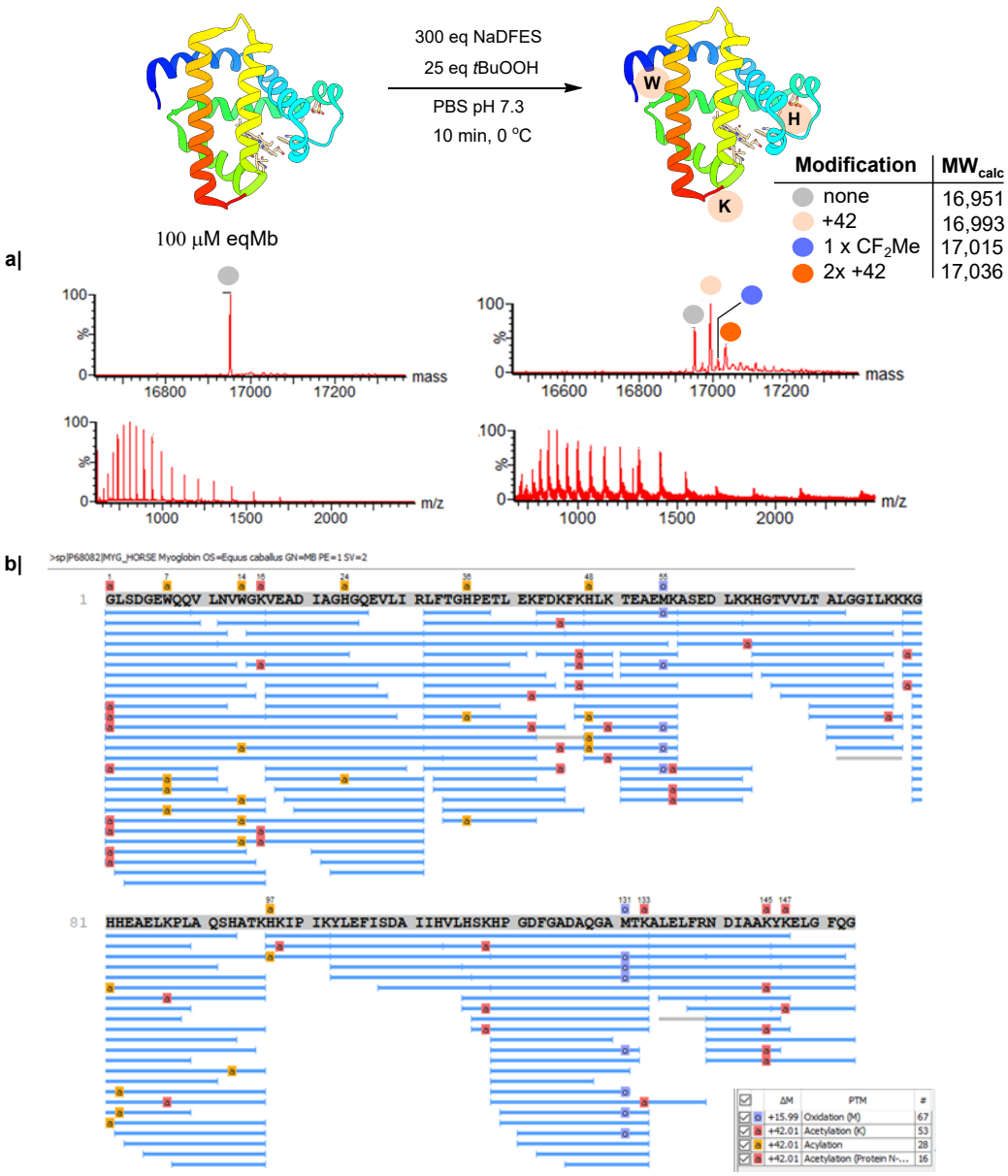


Fig. 4.1: **a**| Modification of eqMb with NaDFES under conditions similar to those optimised for trifluoromethylation, with LC-MS monitoring of the reaction mixture. **b**| Tryptic peptide map (PEAKS software) with annotated confident sites of modification (red - Lys or N-terminus acetylation, yellow - His or Trp, violet - residual Met oxidation).

An obvious candidate for the identity of the $\Delta m = +42$ modification could be deduced from the molecular mass matching that of an acetyl group (-C(O)Me), bearing the structural similarity to the -CF_2Me group.

4.2.2 Elucidating the mechanism

In order to confirm such residue selectivity, a model peptide, H_2N-KFWAWH-COOH, containing relevant residues (Lys, His, Trp) was synthesised and the experiment repeated. Replicating conditions used for the protein modification resulted in the formation of products characterised by sequential $\Delta m = +16$ mass additions, indicating oxidative damage. However, lowering the amount of the oxidant allowed observation of essentially the same acetylation pattern, in addition to low amounts of difluoroethylation - Fig. 4.2a. The modification sites were determined via MS/MS fragmentation and *de novo* spectra reconstruction that confirmed Lys as the most abundant acetylation site, alongside Trp. Additionally, *bis*-acetylation was found within N-terminal Lys, while difluoroethyl group was localised on Trp residues - Appendix (Fig. 4.15). The modified peptides were not resolved via LC, and thus some low abundance modification might not be visible in this statistically probed system.

Control reactions were performed to understand the origin of observed modifications. Surprisingly, N-acetylation was also observed in the absence of the oxidative initiator *t*-BuOOH, where a low amount of acetylated products was exclusively located on the Lys residue as indicated by the MS/MS analysis - Fig. 4.2b. To assess whether the modification originated from a reactive intermediate or NaDFES itself, the non-initiated reaction with peptide was repeated with the reagent pre-incubated for 24 hours and the degree of acetylation compared - Fig. 4.2c. This revealed ∼6-fold faster conversion in the same time, indicating a build-up of an acetylating intermediate during the reagent pre-incubation time.

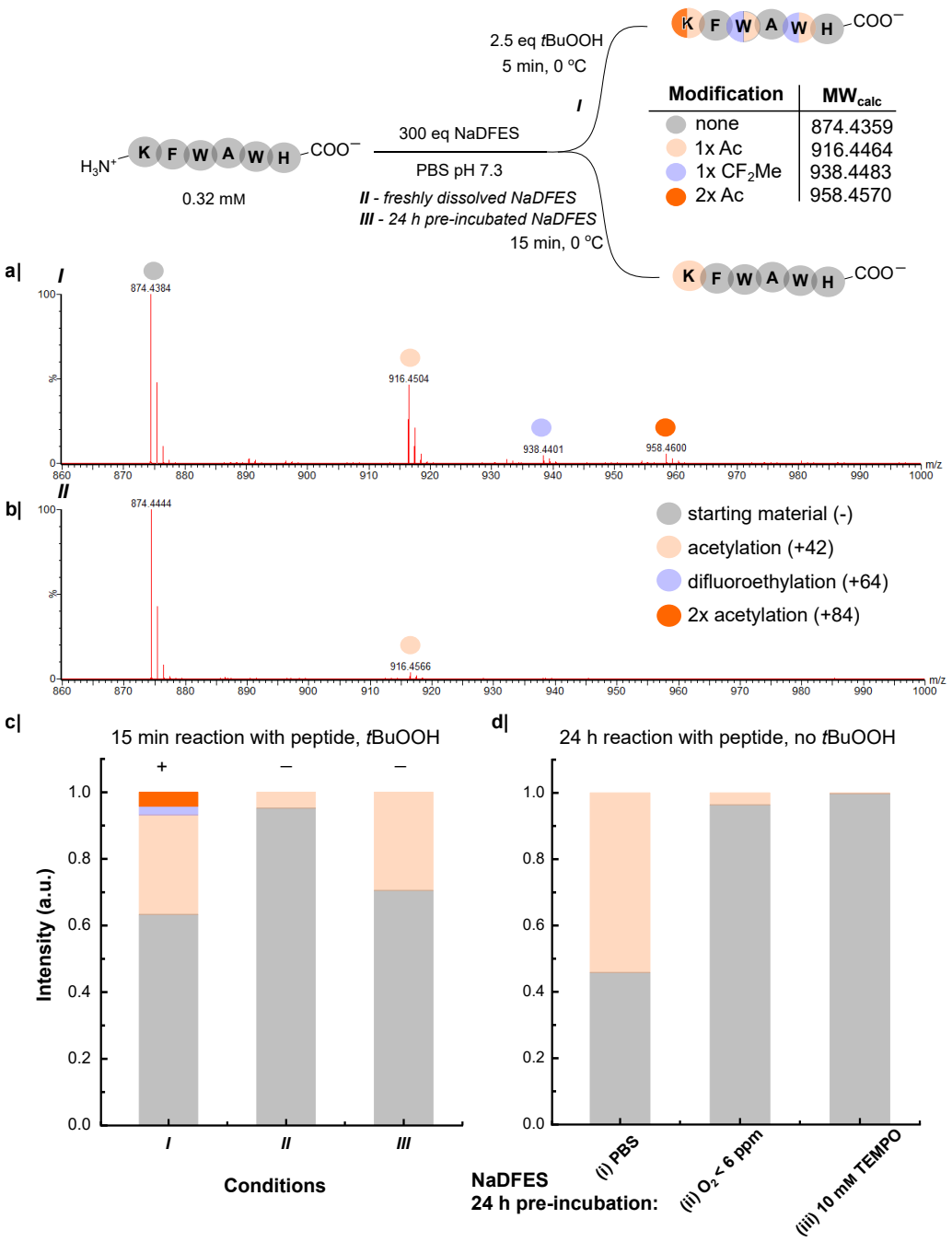


Fig. 4.2: LC-MS analyses of the model peptide modification with NaDFES. Combined (with all peptide products) mass spectrum of the reaction mixture: **a**| in the presence of *t*-BuOOH (conditions ***I***), **b**| in the absence *t*-BuOOH, using freshly dissolved NaDFES (conditions ***II***). **c**| Comparison of LC-MS product distribution after 15 mins of the reaction. Conditions ***III*** used a 24 h pre-incubated NaDFES solution. **d**| Comparison of LC-MS product distribution after 24 h of incubation (no *t*-BuOOH catalysis) with NaDFES solutions pre-incubated under different conditions, (i) open to air PBS, (ii) in an oxygen free environment PBS, (iii) open to air PBS but in the presence of 10 mM persistent radical.

Further exploring the observation for that no modifications were observed for a TEMPOL inhibited protein reaction, the pre-incubation experiment was repeated for a NaDFES solution in the presence of this persistent radical as well as for a stock solution kept in an oxygen free condition (N_2 atmosphere in a glovebox). These were used for peptide modification and the reaction mixtures were analysed with LC-MS after 24 hours -Fig. 4.2d. The analysis revealed that acetylation was substantially inhibited in both of the cases, suggesting that an acetylating intermediate arose from what was potentially a radical mediated oxidative decomposition pathway.

To verify this hypothesis, NaDFES was dissolved in aqueous DMSO and incubated under different conditions. ^{1}H and ^{19}F NMR analyses revealed the formation of new species upon incubation in the presence of air but not in the presence of TEMPOL - Fig. 4.3a. The proton NMR analysis reported on the formation of at least four new species - two of them exhibited similar J_{H-F} coupling patterns, which were postulated to be oxidised sulfinate ($^{-}OSOOCF_2Me$ = oxNaDFES) and an unknown intermediate (Int - possible peroxy derivative)[324]. The other two displayed markedly different appearance, one being a doublet with a characteristic ^{3}J fluorine splitting (AcF), and the other being a singlet corresponding to (AcO^{-}). Additional NMR experiments and extensive literature search allowed unequivocal identification of the species responsible for these peaks as acetyl fluoride and its hydrolysis product (acetate), respectively[325]. Such an assignment was feasible based on comparison of J_{H-F} and J_{C-F} couplings as well as chemical shifts[236] - Fig. 4.3b-c.

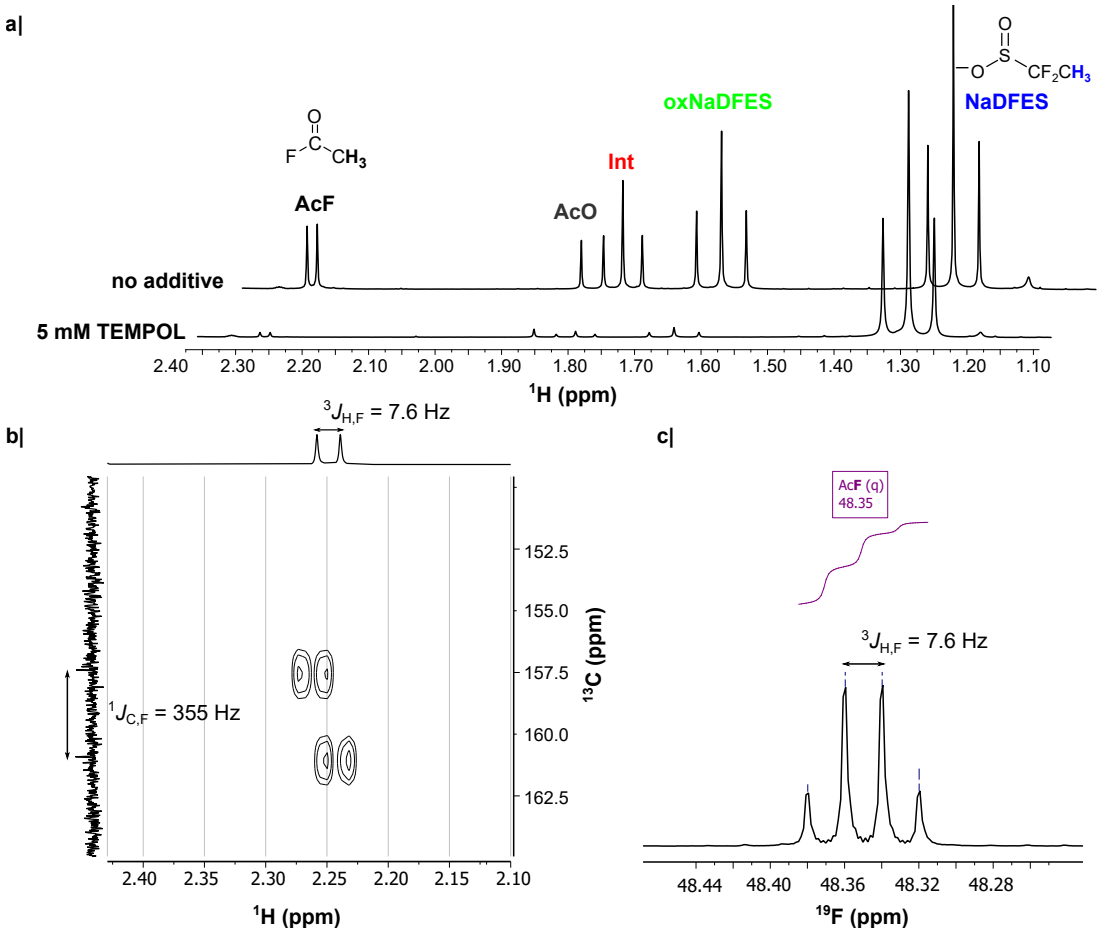


Fig. 4.3: **a**| Fragment of ^{1}H NMR spectrum of 20 mM NaDFES incubated in 'wet' DMSO-d6 (+3 μL D_2O) after 37 hours compared with a spectrum of a sample supplemented with 5 mM persistent radical TEMPOL at the time of dissolution. The peaks were putatively assigned to decomposition products of NaDFES. **b**| Fragment of ^{1}H - ^{13}C HMBC NMR spectrum revealing cross-peak to a characteristic carbonyl carbon with ^{1}J and ^{3}J fluorine coupling constants annotated. **c**| Fragment of the ^{19}F NMR spectrum correlating quartet fluorine signal coupled to the proton signal of interest matching that of AcF [236].

Another major peak present in the ^{19}F NMR was that of HF ($^{1}J_{H,F}$ = 406 Hz, not shown) - in agreement with the loss of basic fluoride needed to generate acetyl fluoride and its hydrolysis product. In the next step, progressive generation of the acetyl fluoride (AcF) was observed in a time-course ^{1}H NMR analysis - Fig. 4.4a.

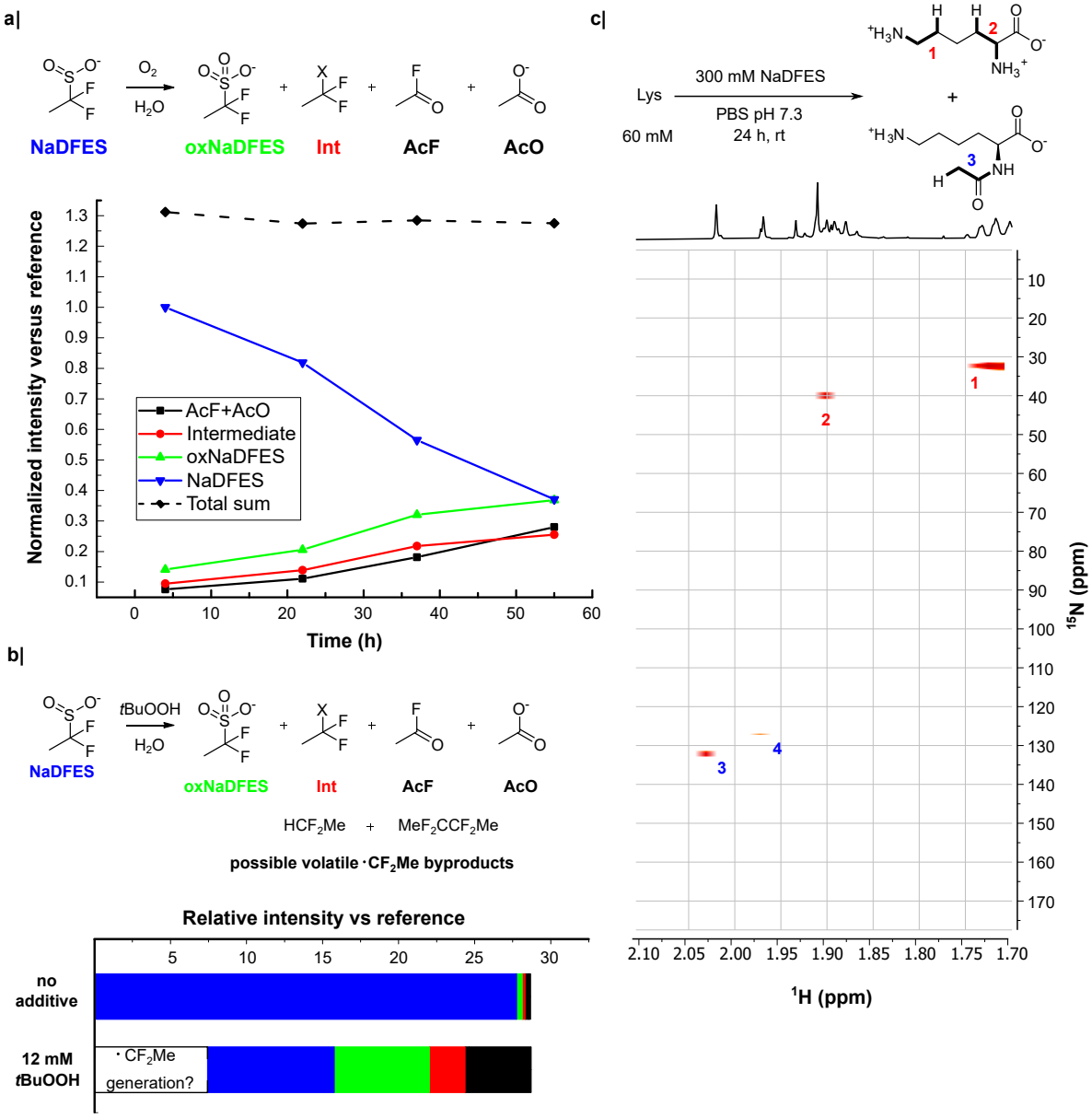


Fig. 4.4: **a-b**| NMR analyses of NaDFES solution under different conditions. (a) Time-distribution of the observed species as quantified by ^{1}H NMR against an internal reference (residual solvent peak). (b) Comparison of product distribution after 1 h incubation of 24 mM NaDFES sample with and without 12 mM of oxidative initiator in DMSO-d6. **c**| Fragment of ^{1}H - ^{15}N HMBC NMR spectrum for a Lys incubated with 300 mM NaDFES revealing 3 bond coupling cross peaks of the starting material (1-2 in red) as well as main N(α) acetylated product (3 in blue). The low intensity signal (4 in blue) most likely stemmed from a lower abundance N(ε)-acetylation product [326–328].

Main species were quantified against an internal reference, which revealed grad-

ual conversion of NaDFES into the its decomposition species (as the total sum of intensities remained constant). Their distribution was compared with that of generated in a NaDFES solution with the oxidative initiator - Fig. 4.4b. As expected, the same species appeared but in much shorter time. A missing mass balance was tentatively assigned to volatile products of radical recombination/hydrogen abstraction. Importantly, under such conditions, around 15% of the NaDFES was converted into acetyl fluoride. AcF is a known acetylating reagent, although usually not used under aqueous conditions[329].

To verify that this decomposition pathway can lead to N-acetylation in the aqueous buffered conditions, NaDFES was incubated with a solution of Lys and the reaction mixture was analysed by NMR spectroscopy. Although such conditions failed to produce significant amounts of product, leaving the majority of the starting material unreacted, N-acetylated products could be detected via ^{1}H - ^{15}N HMBC NMR analysis - Fig. 4.4c. This ruled out the participation of heteroaromatic residues in the reaction mechanism. The low conversion obtained in the reaction could be rationalised by a competing hydrolysis of the acetyl fluoride which, although typically slower when compared with other acyl halides[330], would effectively limit the amount of accessible reagent unless constantly produced in a high excess as in the case of peptide and protein reactions. Altogether, the observed N-acetylation was most likely caused by reactivity of acetyl fluoride generated by decomposition of the NaDFES reagent.

4.2.3 Potential applications in mild protein acylation

Following on from this observation, it was wondered whether the interesting kinetic profile of this N-acetylation would allow a facile discrimination of the N-terminus due to its $\sim$ 2 unit difference in pK_a as compared to other amine groups present in proteins[327]. At neutral pH and an equimolar ratio this should translate to $\sim$100 fold difference in the effective nucleophilic amine concentration, potentially inducing some level of selectivity. First, the results from the small molecule (Lys) experiment were evaluated, for which the pK_a difference is smaller ($\sim$1.5)[331]. ^{1}H NMR quantification of the acetylated products, indicated only 1.3 fold preference for the reaction of the more nucleophilic N(α)-amine.

To further evaluate this hypothesis, the reaction was performed on melittin - a peptide that contains three Lys residues. This substrate was incubated for 24 hours with NaDFES in a buffered solution and the reaction mixture was analysed with LC-MS/MS - Fig. 4.5. Analysis of the combined mass spectra revealed $\sim$55% conversion to the singly modified peptide with a very low amount of double addition - Fig. 4.5a. In the carried out analysis, the different N-acetylated peptides could be separated from the starting material, but not from each other - Fig. 4.5b. Fragmentation analysis confirmed formation of N-acetylated products on both the N-terminus and Lys residues. To determine the relative ratio of different populations, the intensity of fragment ions characteristic to specific products was compared. To account for possible differences in fragmentation efficiency, a fragment common with the non-modified starting material was selected, which allowed normalisation of the signal.

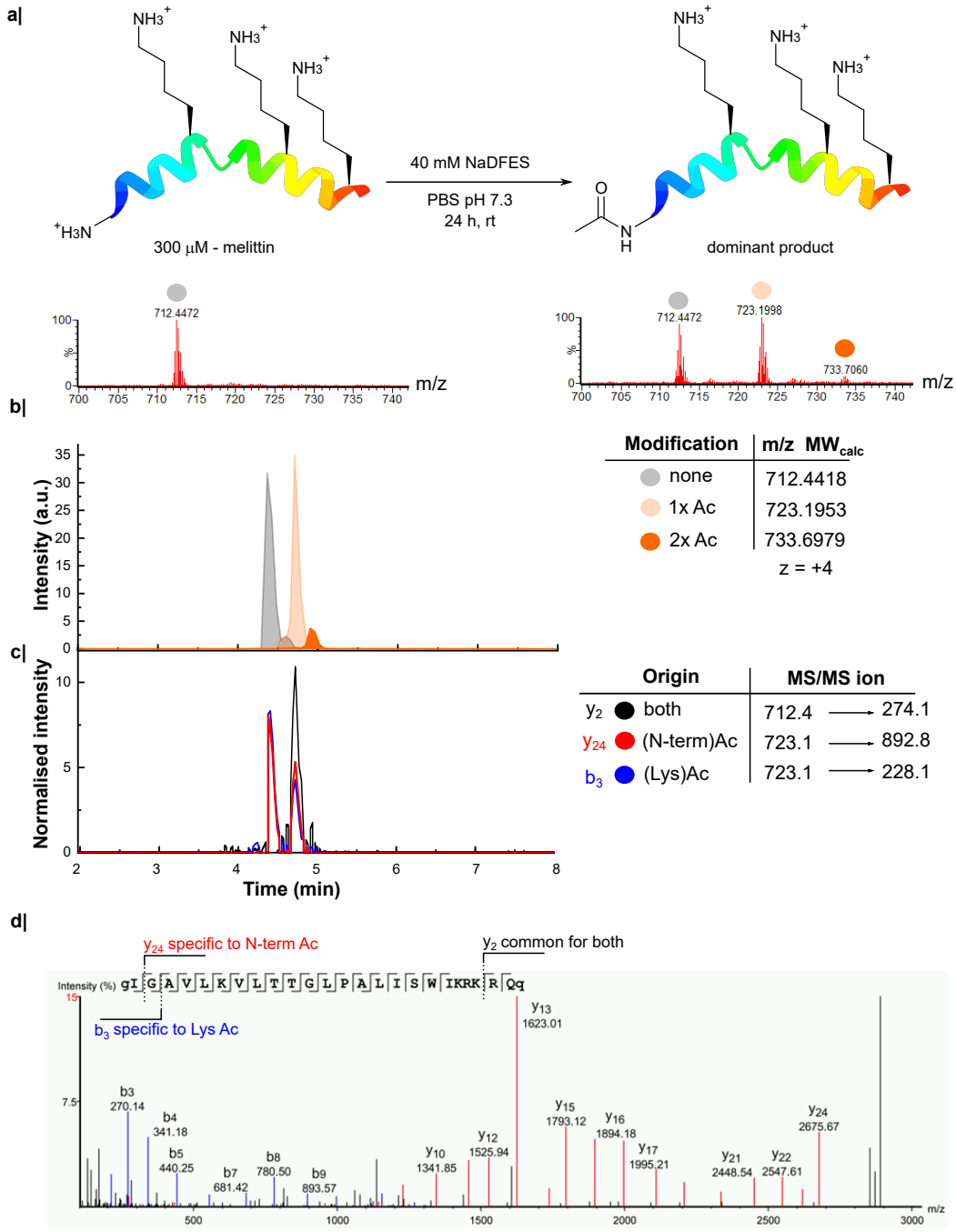


Fig. 4.5: LC-MS/MS analysis of melittin incubated with NaDFES in the presence of air. **a**| Mass spectrum ($z = +4$) of the reaction mixture after 24 h of incubation revealed the formation of expected acetylation products. **b**| Selected ion chromatogram (MS) with starting material, acetylated and double acetylated product peaks extracted. **c**| Selected ion chromatogram (MS/MS) with ions corresponding to fragmentation of any form of melittin (y_2 - black), N-terminus acetylated (y_{24} - red) and any Lys-acetylated (b_3 - blue) extracted. **d**| Melittin N-terminal acetylated product fragmentation spectrum, showing the annotated ions sequence.

In this way, it was estimated that the modification of N-terminus was preferential as compared to any of the possible Lys amino groups, but only in 1.5 fold - Fig. 4.5c-d. This confirmed the weak trend of the N(α) amino group reactivity. Considering typically high Lys content as well as local variation in their pK_a (hyper-reactivity)[142] it was deemed unlikely that useful levels of selectivity could be achieved for a protein reaction.

Study of the reaction on a protein substrate - lysozyme revealed that a NaDFES concentration of 20 mM was optimal, resulting in 0.7 acetylation copy number after 24 h of incubation - Fig. 4.6a. The reaction rate did not increase substantially upon further increase of the reagent excess, which may have resulted from the complete use of the dissolved oxygen and further limited by its diffusion rate from air. Other proteins - myoglobin and biologically relevant acetylation target[332] – histone H2A, could be modified in a similar manner Fig. 4.6b. Finally, it was verified that N-acylation reactivity was not limited to NaDFES, but could also be performed with other difluoroalkyl sulfinates, as use of difluoroheptylazidosulfinate enabled the attachment of an azide, useful for bionconjugation purposes - Fig. 4.6c. Presence of the corresponding acyl fluoride from this compound was also verified via NMR analyses (data not shown).

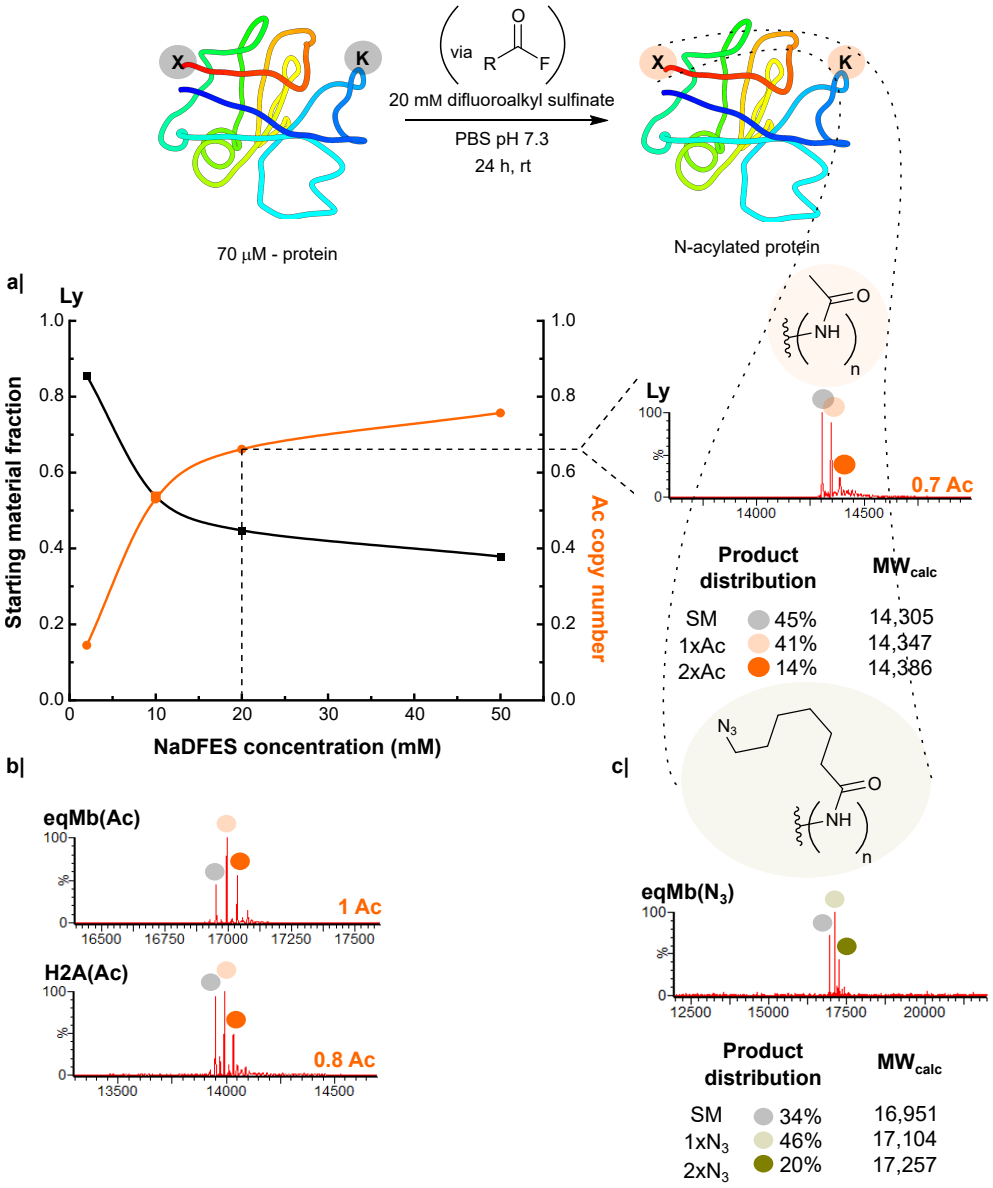


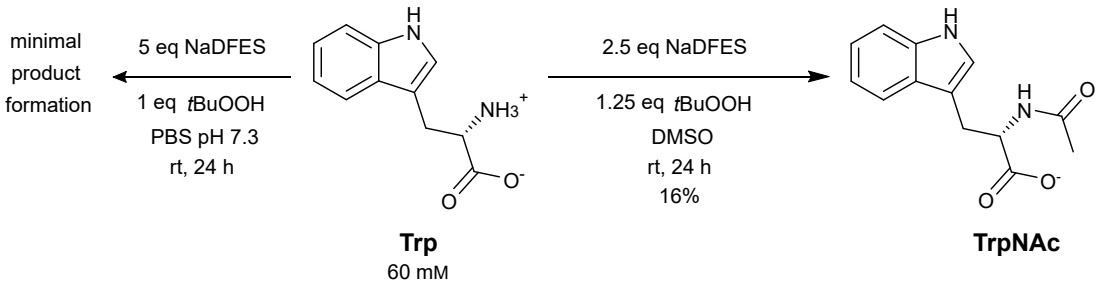
Fig. 4.6: LC-MS analyses of protein acylation with fluoroalkylsulfinates. **a**| Quantification of lysozyme acetylation versus NaDFES concentration after 24 h of the reaction. Mass spectrum of the reaction mixture incubated with 20 mM NaDFES. **b**| Mass spectra of acetylation of different proteins: myoglobin (eqMb) as well histone H2A (H2A) together with the calculated modification copy number. **c**| LC-MS of eqMb incubated with 20 mM difluoroheptylazidosulfinate after 24 h of the reaction confirmed the formation of expected N-acylated products ($\Delta m = +153$).

4.3 NaDFES as a C-acetylating agent

Having proposed a plausible mechanism for N-acetylation, next experiments were aimed at the investigation of the apparent C-acetylation of heteroaromatic residues.

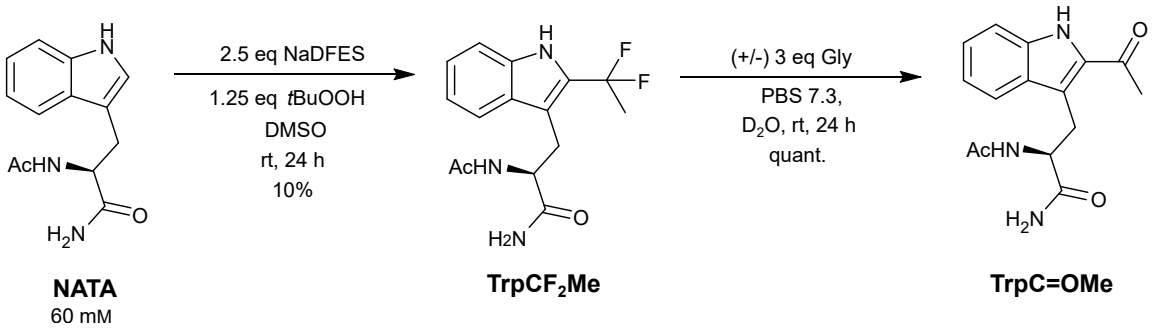
Gem-difluoro motifs attached to electron-rich rings (here indole[333] and imidazole[334] moieties - present in Trp and His respectively) are known to be prone to hydrolytical loss of fluorine under aqueous conditions yielding acyl derivatives[335]. Such reactivity would readily explain the observed modification of both Trp and His, separating it from the N-modification pathway.

In order to verify this hypothesis, a small molecule model was employed. Trp was reacted in an aqueous buffer under an excess of reagents and the reaction mixture analysed by NMR. ^{1}H NMR indicated only a minimal consumption ($< 5\%$) of the starting material after 24 h of the reaction. Additionally, ^{19}F NMR did not confirm formation of new fluorinated products (data not shown). Further increase of reagent excess and/or temperature resulted in the formation of complex mixtures that could not be resolved chromatographically. Next, the reaction was carried out using DMSO as a solvent which, instead of the fluorinated product, allowed the isolation of an N-acetylated derivative **TrpNAc** in a low yield, together with unreacted Trp - Sch. 4.1.



Sch. 4.1: Difluoroethylation of Trp in different solvents (0.12 mmol).

To simplify the analysis, the reaction was repeated with the N-protected substrate - NATA (N-acetyltryptophanamide). This indeed allowed isolation of a single difluoroethylated derivative as the main product, though in minimal yield. To test the hydrolysis hypothesis, the purified **TrpCF$_2$Me** product was incubated in an aqueous buffer solution. This led to a rapid elimination of fluoride, resulting in a quantitative conversion to the ketone product **TrpC=OMe**. Incubation of **TrpCF$_2$Me** in the presence of Gly, with a free amine group, did not lead to a different result, confirming separation of the N- and C- acylating reaction pathways - Sch. 4.2.



Sch. 4.2: Difluoroethylation of NATA and subsequent hydrolysis in deuterated buffer.

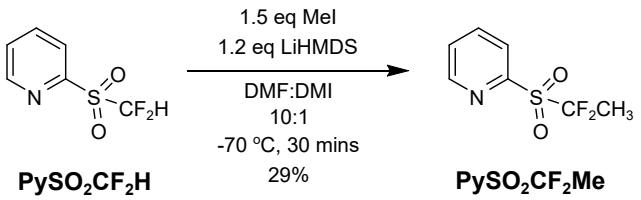
A similar difluoroethylation experiment for HisNAc did not yield conclusive results, as, in addition to low conversions, formation of many products was observed, which led to complex and unresolvable mixtures. This suggested that difluoroethylation might have led to formation of both possible regioisomers, as acidic conditions were not used[78]. In the view of low reactivity as well as the formation of side products, further investigation of difluoroethylation using fluoroalkylsulfinates was not pursued.

4.4 Alternative $\cdot CF_2CH_3$ generation system

To continue the studies, it was deemed necessary to avoid the identified off-pathway reactivity. Under the proposed protocol, N-acylation was intrinsically connected with the desired difluoroethylation due to the instability of the precursor compound. In addition, overall low conversions to the fluorinated products suggested poor efficiency of the radical initiation. Therefore, alternative systems capable of biocompatible generation of the $\cdot CF_2CH_3$ from a stable, yet reactive precursor were explored.

4.4.1 Photoredox activation of difluoroethyl pyridyl sulfone

First, structurally similar scaffolds were taken into consideration. Being formal precursors to sulfinates, (hetero)aryl sulfones are also known to be suitable radical precursors. They can be initiated under various reductive manifolds enabling a plethora of useful small molecule transformations[336,337]. One method that stood out in terms of potential applicability to biomolecule modification, was photoredox catalysis employing the commonly used water soluble $Ru(bpy)_3^{2+}$ catalyst[177,338]. Of the possible heteroaromatic platforms, 2-pyridyl sulfone (Fig. 1.15a) was hypothesised to possess both suitable solubility and stability. To test the suitability for protein difluoroethylation, the desired radical precursor **$PySO_2CF_2Me$** was synthesised in one step from a commercially available compound $PySO_2CF_2H$ - Sch. 4.3.



Sch. 4.3: Synthesis of the **$PySO_2CF_2Me$** radical precursor

Using **$PySO_2CF_2Me$**, the reaction was tested for the model peptide under a range of conditions. In order to produce the radical, Ru(I) had to be accessed using a sacrificial reductant[177]. Ascorbic acid is a known biological reductant, capable of reducing excited states of photocatalysts[181]. Moreover, $Ru(bpy)_3^{2+}$/ascorbate system in the presence of proteins was found to be highly functional and biocompatible[339]. Other compounds did not allow the desired difluoroethylation reaction to be observed (data not shown). The control experiments verified the necessity of all elements of the system (catalyst, light, ascorbate). In the absence of ascorbate, extensive oxidative damage was observed (evidenced by a characteristic mass spectrum) - Fig. 4.7a. Therefore, the presence of ascorbic acid also significantly reduced the oxidative damage, as evidenced by the reduced intensity of $\Delta m = +32$ ladder products - Fig. 4.7b-d.

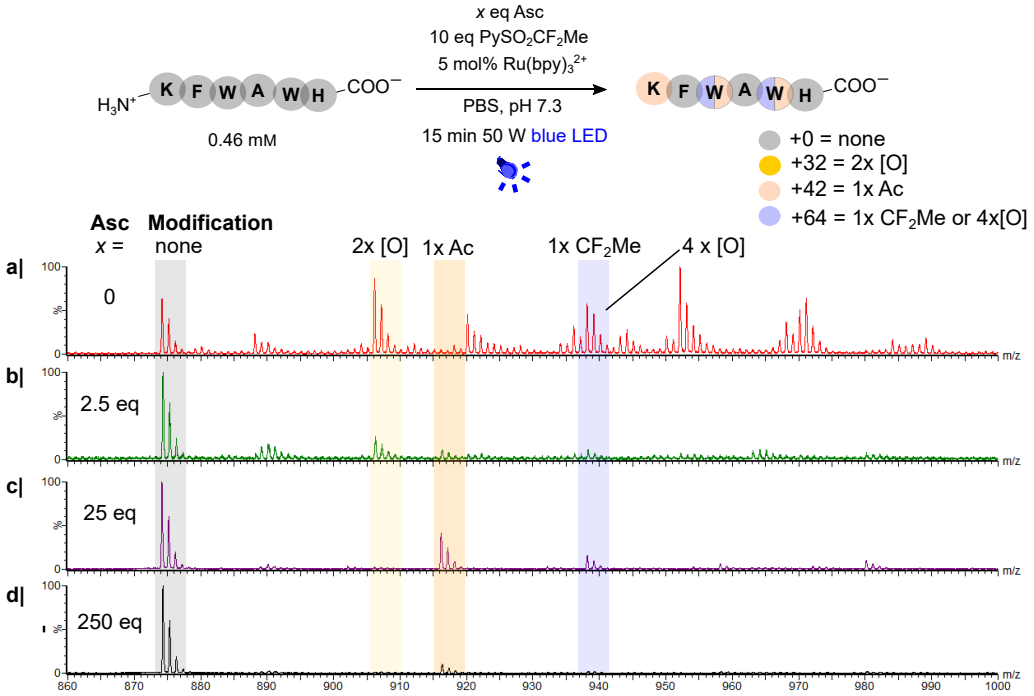


Fig. 4.7: LC-MS/MS screening of the difluoroethylation of the model peptide under photoredox conditions. **a-d**| Combined mass spectrum of the reaction mixture (a) in the absence of ascorbate (Asc) as a reductant and (b-d) with different amounts of ascorbate.

Screening reaction mixtures via LC-MS/MS showed that for an optimal conversion to the expected products, the excess of radical precursor and reductant should be similar Fig. 4.7c-d. The low amount of the difluoroethylated species suggested rapid hydrolysis of this motif in aqueous buffers, which was not investigated in detail at this stage. LC-MS/MS analysis of the acetylated species revealed that the most abundant modification site was Trp. Additionally, a minor (lower intensity) population of modification was localised on the N-terminus (Lys) - Fig. 4.8. The two products could be separated chromatographically, and their identity established based on their fragmentation pattern. This suggested that changing the radical generation system only partially eliminated the undesired N-reactivity.

In order to confirm this observation, a chemical assay differentiating between the N- and C-acetylation was developed. It was based on the divergent, intrinsic reactivity of the new acyl moieties. The two products could be unequivocally distinguished upon the reaction with an excess of sodium borodeuteride, which would exhaustively reduce the C-acetylated product - ketone, to the corresponding alcohol, while leaving the N-modified product - amide, intact[340]. The assay revealed that Trp hydrolysis products disappeared upon the reduction, while the Lys localised remained intact - Fig. 4.8. Furthermore, the presence of the reduced species could be detected, further confirming the envisioned reaction course.

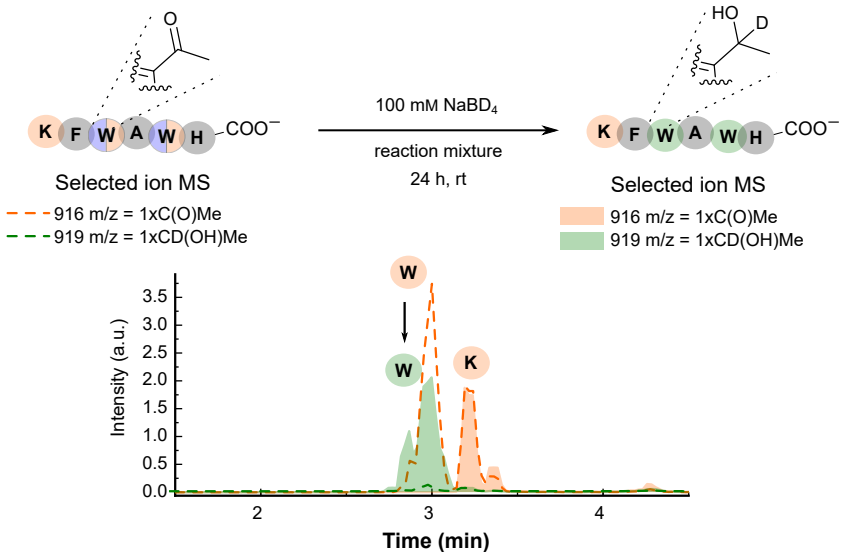


Fig. 4.8: LC-MS analysis of difluoroethylated model peptide subjected to the reduction assay. The graph shows fragment of the extracted ions chromatogram (acetylated - peach, and reduced products - green) of the reaction mixture before the addition of the reducing agent (dashed lines) as well as after 24 h incubation (shaded shape).

To investigate the possible origin of such reactivity, the reaction mixture containing the **PySO$_2$CF$_2$Me** was investigated by NMR spectroscopy before and after the irradiation with blue light - Fig. 4.9.

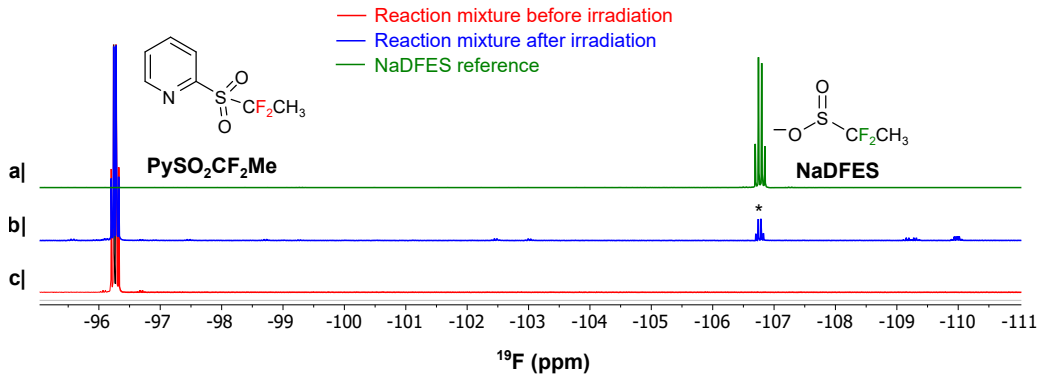


Fig. 4.9: ^{19}F NMR spectra of **a**| 10 mM NaDFES sample, **b**| mixture of **PySO$_2$CF$_2$Me** radical precursor irradiated with blue light as in the difluoroethylation reaction in aqueous buffer, **c**| and the same solution before the irradiation.

Although the compound was stable in the reaction mixture when kept in the dark for

as long as 7 days, multiple new species were formed when irradiated. Interestingly, one of them could be identified as NaDFES, which had been previously shown to be responsible for the acetylation reactivity. In situ generation of this compound and possible catalysis (via photooxidation) of its further decomposition into the acetyl fluoride could most likely account for the observed residual N-modified side products.

4.4.2 Investigating protein difluoroethylation

Considering a rather scarce generation of NaDFES (and therefore AcF) within the reaction time, it was hypothesised that the undesired reactivity could be overcomed by the addition of a sacrificial amine into the reaction mixture. A natural candidate was Lys, as a direct and equivalent competitor to protein residues. This was deemed to be suitable since Lys in high concentrations is typically used as a buffering agent (up to hundreds mM). The borodeuteride assay was again employed for the evaluation of this hypothesis - Fig. 4.10. In this way, it was possible to determine that in a solution containing 100 mM Lys, the N-acetylation of the peptide was effectively reduced by >90% based on comparison of single acetylated products (assuming complete reduction of the heteroaromatic ketone). Unfortunately, this assay did not yield semi-quantitative information about the Trp vs His selectivity of the reaction since different products were not resolved chromatographically. MS/MS fragmentation indicated single instances of the C-modification localised within His residues, suggesting low chemoselectivity (as Trp residues were of an effectively double concentration).

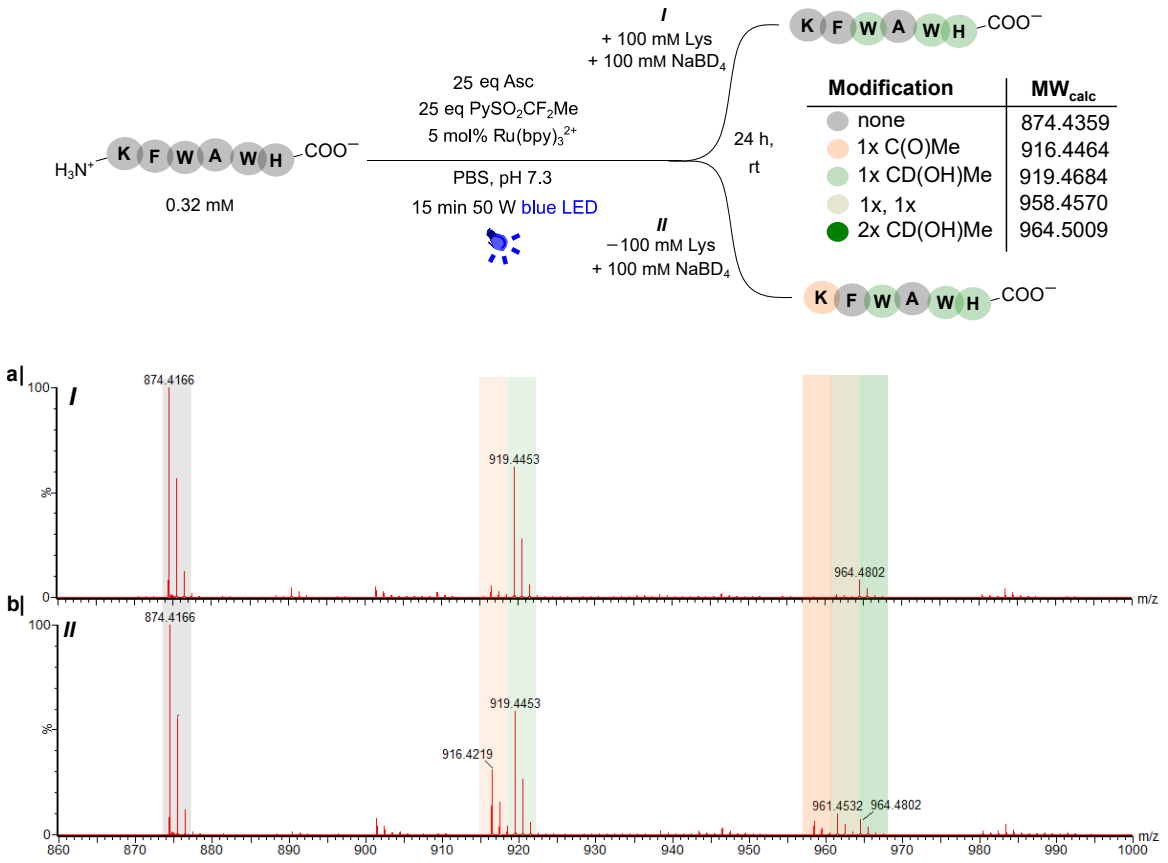
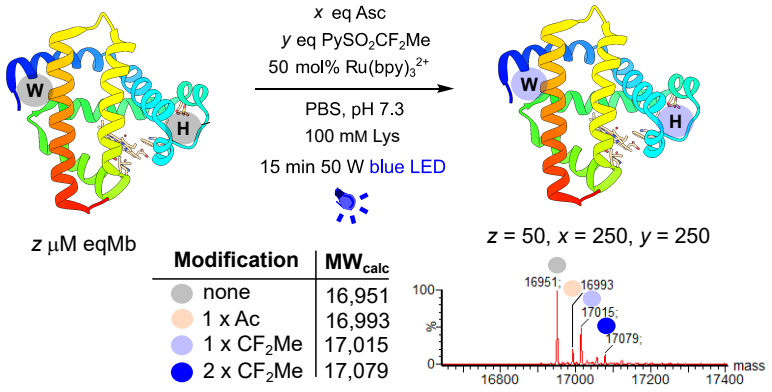


Fig. 4.10: LC-MS analysis of the borodeuteride reduction assay for the difluoroethylated model peptide. **a-b**| Combined mass spectra of the reaction mixture: (a) incubated in the presence of 100 mM Lys, (b) without the amine.

Interestingly, an increase of the reagent excess and/or irradiation time, did not result in a significant increase observed conversion (<50%), suggesting low efficiency of the reaction and possible degradation of the peptidic substrate.

Having established conditions in which almost exclusive difluoroethylation was possible, it was wondered whether the photoredox protocol could be applied to protein substrates. To this end, various conditions were tested on a model protein myoglobin and reaction mixtures were analysed by LC-MS - Tab. 4.1. In order to observe significant conversions, it was necessary to employ high reagent excess (>250 eq).

However, it was found that such a high concentration of reagents affected protein stability as protein precipitation was readily observed.



Tab. 4.1: LC-MS screening of photoredox difluoroethylation of a model protein - myoglobin. In the figure spectrum for entry 4. In some samples, extensive precipitation (prec.) could be observed. The calculated copy number (copy nr.) refers to the average number of the CF_2Me modification.

Entry	**(protein conc.)** z [μM]	**(Asc)** x [eq]	**(PySO$_2$CF$_2$Me)** y [eq]	**Conversion copy nr.**	**Prec.**
1	50	10	10	< 0.05	no
2	50	50	50	0.28	no
3[a]	50	50	50	-[b]	yes
4	50	250	250	0.43	yes
5[a]	50	250	250	-[b]	yes
6[a]	50	250	100	0.16	no
7	100	250	250	0.63	yes
8	200	250	250	0.65	yes
9[c]	100	250	250	0.57[d]	no

[a] - irradiation time = 45 min, [b] - ion series did not deconvolute, [c] - with 6 M urea, [d] - together with much higher acetylated product copy number = 0.3.

Similarly, prolonging irradiation time also led to the protein degradation as evidenced by complex mass spectra. Unfortunately, using lower amount of reagents or altering the stoichiometry produced only minimal conversion to products. Other pro-

tein substrates (annexin V and SRP19 - not shown) gave similar results, suggesting low reactivity of the radical and/or unsuitability of the photoredox conditions. The problems with the stability of the protein could be circumvented by the use of denaturing conditions (high urea concentration - Tab. 4.1 - entry 9). In this case, moderate conversions to the difluoroethylated protein could be observed, accompanied with a higher proportion of the hydrolysed product. Interestingly, various difluoroethylation conditions of folded myoglobin resulted in only negligible amounts of ketone product being detected (<10%).

This observation suggested that hydrolysis of the fluoroalkylated motif to the corresponding ketone might be inhibited for folded proteins, in which the modification is localised within solvent inaccessible aromatic residues - Tab. 2.2. To probe this hypothesis, samples under native and denaturing conditions (Tab. 4.1 - entries 7 and 9) were re-analysed after 24 h to assess the extent of hydrolysis - Fig. 4.11. This indeed revealed that for the folded sample, the distribution of products did not change significantly, while for the sample incubated in the presence of denaturant, the amount of fluorinated products decreased in favour of the hydrolysed ones.

Considering difficulties in achieving good conversion for protein substrates, apparent low chemoselectivity and further complications by impaired hydrolysis of the fluorine tag (perhaps localised within aromatic systems of Phe and Tyr which did not promote the transformation), the studies on protein difluoroethylation were concluded at this point. Investigations for different radicals and further studies on the photoredox system will be described in the next chapter.

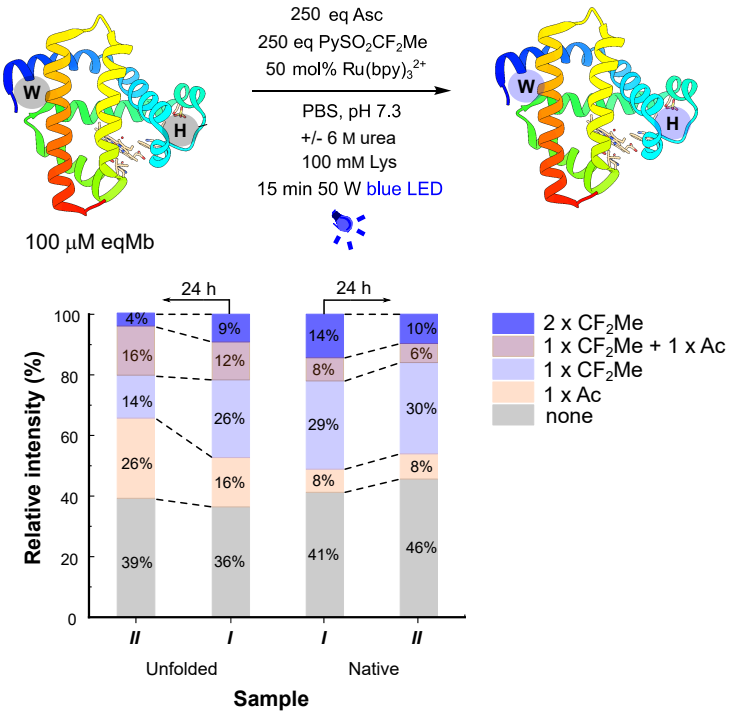


Fig. 4.11: LC-MS analysis of the product distribution immediately after the difluoroethylation reaction (***I***) and after 24 h incubation in the reaction mixture (***II***), under native or denaturing conditions.

4.5 Conclusions and discussion

In this chapter, following on from the trifluoromethylation work, further fluoroalkylation methodologies for the modification of native residues in proteins were investigated, using difluoroethylation as an initial model system. Utilisation of a system analogous to that used for trifluoromethylation, based on sulfinates as radical precursors, did not proceed as expected. Screening of reaction conditions using protein and peptide substrates revealed a series of both N- and C-acetylated products being formed, once again reinforcing the challenge of achieving selectivity in the sea of functional groups present in proteins[341].

Careful studies of reagents and reactions with small molecule and peptide substrates allowed elucidation of the mechanism of the observed N-acetylation. It was proposed that difluoroalkyl sulfinates oxidatively degrade to acylating agents - acyl fluorides. Although the sulfinic acids are well known for their instability due to disproportionation[342] and the autooxidation of sulfinic acids in aqueous solution has been studied in detail[324], to the best knowledge of the author there are no reports regarding the decomposition of difluoroalkyl sulfinates in aqueous buffers of neutral pH. However, a similar pathway of decomposition was observed for other fluoroalkyl radical precursors when investigated in aqueous conditions[343]. Here, this serendipitous observation was demonstrated to be useful in a novel, extremely mild N-acylation of proteins.

Since acetylation (N-termini and/or ε-amine group of lysine) is one of the most important post-translational modifications of many proteins (being implicated in essential biological processes e.g. cell signalling and metabolism)[344] one can predict that this chemical acetylation protocol would be of interest to the community, offering an alternative to the widely used acetic anhydride acetylation method[345]. There have been a number of studies which involve the in vitro chemical acetylation of proteins in order to understand the role of organisms' acetylome and its impact on enzymes' activity[346]. In some cases acylation is used for derivatisation prior to MS proteomics experiments[347]. For such applications this method could be advantageous, as gradual generation of acetylating agent removes the need for a protein dependent optimisation of reaction conditions using hazardous acetic anhydride[348].

Furthermore, the hydrolysis of the acetyl fluoride is slow in neutral aqueous

buffer[349], which also facilitates control of the reaction mixture pH, crucial to avoid side reactions. For applications requiring faster kinetics, the rate of acetyl fluoride formation could be accelerated in the presence of an oxidative activator. On the other hand, direct use of acetyl fluoride for this type of reaction may be inconvenient and potentially hazardous due to its low boiling point (∼21 °C)[331]. Other acyl halides are either unstable (iodide)[350] or are characterised by a much higher rate of hydrolysis (chloride, bromide)[351] as compared with acetyl fluoride. Finally, different specialised but commercially available acylating reagents (such as N-acylazoles) that can be used in aqueous solutions, often have preferential reactivity with O-nucleophiles[352,353]. Apart from the established use of acyl fluorides in peptide synthesis[329], it is worth noting that their desirable reactivity profile for the modification of biomolecules has been successfully exploited for preparing drug-antibody conjugates[354]. Therefore, this method may find use in general applications for chemical biomolecule acetylations. Interestingly, a similar kind of reactivity was not observed for non-fluorinated sulfinates (*n*-propyl, benzyl), as no mass shift was detected for the protein incubated with such reagents, highlighting the enabling role of the fluorine atoms (data not shown).

Due to the fact that the sulfinate system turned out to be poorly suited for the desired fluoroalkylation, alternative radical generation systems were investigated. This led to the adaptation of structurally similar pyridyl sulfones as radical precursors, efficiently activated using photoredox catalysis. Although this methodology did not allow full decoupling from the other reaction pathways, it was improved enough that undesired N-reactivity could be drastically reduced by the addition of a sacrificial amine to the reaction mixture. This allowed further studies on fluoroalkylation,

which suggested that •CF_2Me may possess much lower chemoselectivity than the CF_3 counterpart. More importantly, it did not allow for an efficient product formation on the folded model protein, also due to biomolecule stability issues under the reaction conditions.

With the above in mind, as well as the fact that the studies on the CF_2Me radical were to establish suitability of fluoroalkylation for bioconjugation (with a general group R), it was deemed unlikely that such a strategy may provide enough advantages over existing methods. Supporting this hypothesis were recent reports in the literature, where reactivity of various radicals have been investigated on peptide and protein models. It was clearly demonstrated that protonated His residues are good acceptors of alkyl radicals that are nucleophilic in their nature[75]. Similar selectivity was also observed for certain fluoroalkyl radicals, enhancing the notion of the critical protonation step for achieving exclusive His modification[78]. Such behaviour, where the nucleophilicity of the radical is significantly more pronounced for a protonated substrate in acidic mixtures was observed for the radical modification of pyridines[355]. Despite indicating a clear trend, such an approach cannot be readily applied to proteins, as low pH is often incompatible with their native structure and function. On the other hand, working under near neutral conditions, the influence of the alkyl group hinders the ability for chemoselective radical additions to any of the native (hetero)aromatic residues. As such, it seems likely that the relative electrophilicity exerted by the presence of fluorine atoms closely competes with the nucleophilicity inherent to alkyl substituents[155]. This hypothesis was supported by the work, in which full Trp selectivity for •CF_2CF_2R was demonstrated[123]. The extension of the

fluorinated chain ensured sufficient electrophilic character, and allowed for the development of a bionconjugation strategy based on this chemistry, also using photoredox catalysis for the radical generation[356].

One key discovery was that the difluoroethylation of (hetero)aromatics natively present in proteins turned out to yield, not entirely unexpectedly, hydrolysed acyl products. This observation opens a new avenue of investigation as ketones/aldehydes are sought after bioorthogonal handles, often appended to biomolecules due to the variety of bioconjugation chemistries they enable[357]. The observation was fully validated on a small molecule synthetic model for Trp, with other analytical results pointing towards the same behaviour for His residues, also supported by recent reports[78]. An interesting result suggested the hydrolysis kinetics for the difluoroethylated protein may be controlled by solvent accessibility of the modified residues, providing further evidence that $\bullet CF_2Me$ can access residues buried in the hydrophobic core. However, in view of the discussed selectivity issues, it was clear that only radicals with more clearly defined electronics could provide sufficient Trp selectivity and further research on difluoroalkylation was not pursued. A straightforward candidate with such characteristics was the difluoromethyl radical, often resulting in similar reactivity to the trifluoromethyl counterpart[184]. The hydrolysis of this group would result in perhaps the most useful acyl handle - an aldehyde[32]. The development of this idea is described in the next chapter.

4.6 Experimental

4.6.1 General information

For general experimental information please refer to the relevant sections in Chapter 2. Details that are different and specific to experiments described in this and next chapters are given below. The reaction mixtures during 450 nm LED light irradiation were not stirred, while overnight incubations were shaken at 300 rpm. For flash column chromatography CombiFlash NextGen 100 automated chromatography system was used (Teledyne ISCO) with RediSep Rf Gold Silica Flash Chromatography Columns (4 – 20 g of silica, Teledyne ISCO). A sample of histone H2A (*Homo sapiens*, H2A2A_HUMAN) was a kind gift from Dr Ritu Raj (Ben Davis group), the protein was expressed according a known procedure[358]. $[Ru(bpy)_3]^{2+}$ refers to the $[Ru(bpy)_3]Cl_2 \cdot 6H_2O$ (tris(bipyridine)ruthenium(II) chloride) photocatalyst.

Mass Spectrometry

For all mass spectrometry analyses Waters Xevo G2-S QToF MS coupled to a ACQUITY UPLC H-Class LC system (Waters Corporation) was used. Instrument was auto-calibrated throughout analysis by an internal lock-spray mass calibrant (LeuEnk). For protein analyses, desalting was achieved with the ProSwift^{TM} RP-1S HPLC column (250 x 4.6 mm x 5 μm, Thermo Scientific^{TM}) using water (solvent A) and acetonitrile (solvent B), both with 0.1% formic acid, as the mobile phases. The LC program for an intact protein analysis was realised as followed: 95% A $\rightarrow$ (0 – 2) min, gradient to 95% B $\rightarrow$ (2 – 8) min and 95% A $\rightarrow$ (8.1 – 10 min) at a 0.3 mL/min flowrate. Following ESI ionisation source parameters were used: capillary

voltage 3 kV, cone voltage 20 V, source temperature 100 °C, desolvation temperature 400 °C and desolvation gas (nitrogen) flow 700 L/h, cone gas flow 30 L/h. MS was working at 30,000 (FWHM) resolution. For peptide analyses the same LC conditions were used but with a different column (C18 BEH Agilent, 130Å, 1.7 μm, 2.1 mm x 50 mm).

Model peptide tandem mass spectrometry

For peptide reactions, the mass spectrometer operated in data dependent acquisition (DDA) mode with 0.1 s MS survey scan followed by (0.1 – 0.5) s MS/MS scans on the three most intense ions. To induce fragmentation of a broad spectrum of products MS/MS collision energy ramp (15 – 60) V was used. For peptide mapping, the mass spectrometer operated in DDA mode with 0.1 s MS survey scan followed by 0.2 s MS/MS scans on the six most intense ions with mass dependent collision energy. MS/MS data analysis was performed using PEAKS software v8.5 (Bioinformatics Solutions Inc.) using *de novo* reconstruction. Search parameters were set as following: precursor mass tolerance of 10 ppm and a fragment ion tolerance of $\Delta m = \pm 0.02$ Da. No information about peptide sequence was fed into the software and all assigned spectra were manually validated.

Photoreactor design

The photoreactor was assembled in-house (University of Oxford, Electrical Workshops), from commercially available components (Luxeon). Briefly, seven LEDs (LUXEON Rebel ES Royal-Blue 447.5 nm, Luxeon Star Inc, SP-02-V4, 7210 mW @ 700 mA maximum power) were connected in series, and arranged on a round 7-Up

Base mounted on a 60 mm round x 45 mm high heat sink (Luxeon Star, CN60-45B). Power was supplied through a 700 mA dimmable FlexBlock DC driver (Luxeon A011-D-V-700). These were wired with an on/off switch and placed inside of a 305 mm cubic metal box (RS Company). All reactions were performed at maximum available power (50 W).

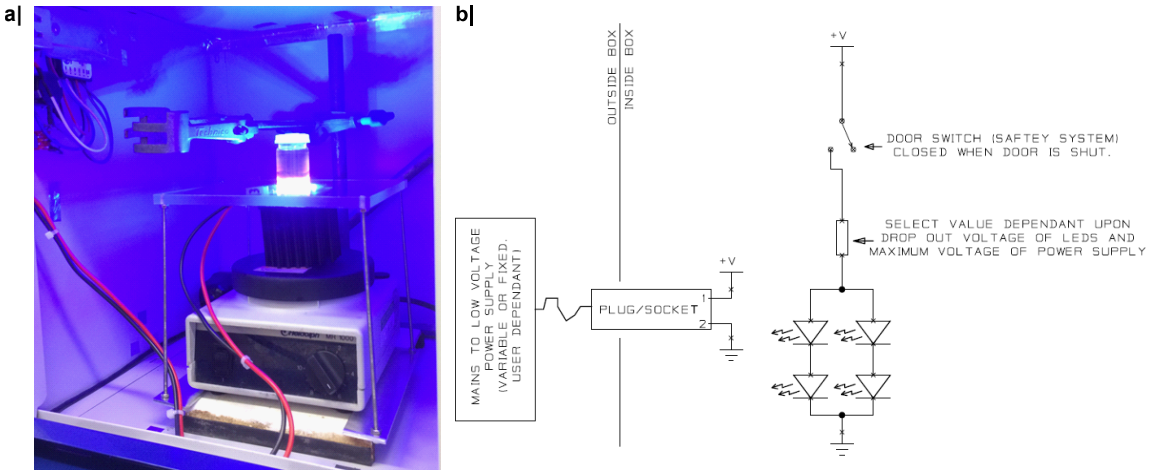


Fig. 4.12: **a|** Picture of the photobox setup with a glass vial in a usual reaction mixture position on a transparent holder. **b|** An electrical diagram illustrating design of the used photoreactor.

4.6.2 Experimental procedures

eqMb reaction with NaDFES and tryptic digest

To a solution of myoglobin (100 μL, 1.8 mg/mL in PBS buffer) kept on ice, sodium 1,1-difluoroethanesulfinate (NaDFES, 4.7 μL, 100 mg/mL in PBS, 300 eq) and *t*BuOOH (3.6 μL, 72 mM in H_2O, 25 eq) were added. The reaction mixture was incubated for 10 mins on ice and after that, an aliquot (1 μL) was withdrawn and diluted with water (60 μL) for LC-MS analysis.

To the remaining reaction mixture, water (31 μL) was added and it was buffer

exchanged with a MiniTrap G25 into 50 mM NH_4HCO_3. To this solution, trypsin was added (3.5 μL, 1:25 w/w, 2 mg/mL) and the mixture was incubated overnight at 37 °C. After that, the tryptic peptides were bound on a ZipTip, washed, eluted (according to the manufacturer's instructions) and analysed via LC-MS/MS.

Model peptide synthesis

Model peptide H_2N-KFWAWH-COOH was synthesised using SPPS following a standard Fmoc-strategy with a pre-attached Fmoc-L-His(Trt)-Wang resin (0.5 mmol/g), using an Activotec peptide synthesiser. The synthesis started with deprotection of the resin (1 g, 0.5 mmol) before the first coupling (25% piperidine in DMF). All processes were monitored with UV. All coupling steps used an excess of reagent solutions: Fmoc-Xaa-OH (5 eq), HBTU (5 eq, 0.48 M), DIEA (6 eq, 1 M) in DMF that were reacted with the resin-bound free amine peptide for 15 mins in a microwave reactor. After the procedure successfully finished, the resin was washed with DCM and dried. The peptide was cleaved off the resin and deprotected in TFA/water mixture/1,2-ethanedithiol (EDT) (92.5/5/2.5, v/v/v, 16 mL) in an ice bath. After 30 mins, the ice bath was removed and the reaction was stirred at rt for an additional 2.5 hours when an LC-UV analysis indicated completion of the reaction. The reaction mixture was added to chilled diethyl ether in a dropwise manner. The white suspension was divided into 50 mL falcon tubes and centrifuged (5 mins, 6000 xg). The procedure was repeated to wash the solid with a fresh portion of diethyl ether. The precipitate was dissolved in H_2O and lyophilised (repeated 3x). This gave KFWAWH as a white powder (372 mg, 0.32 mmol, 64%). LC-UV-MS analysis confirmed > 95% purity of

the peptide. The sequence was confirmed by MS/MS analysis (see elsewhere) and peptide was used without further purification as TFA salt.

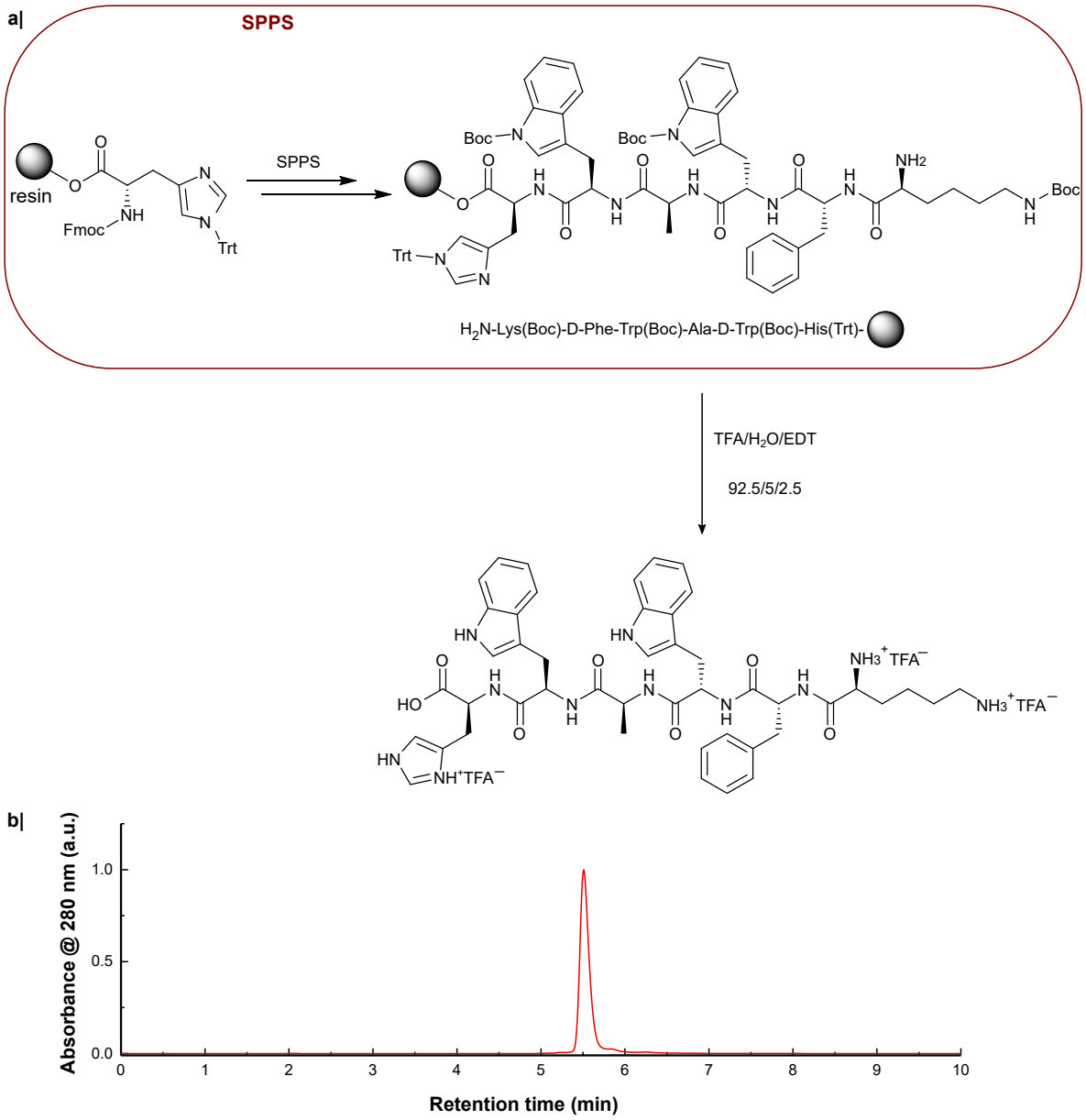


Fig. 4.13: **a**| Scheme of the KFWAWH model peptide synthesis. **b**| LC-UV chromatogram of the product (C18 Gemini NX (Phenomenex)), linear 10 mins gradient of MeCN in aqueous 0.1% formic acid.

Model peptide reactions with NaDFES

NaDFES (15.2 mg/mL, 100 mM) was prepared as a solution in 10x PBS buffer (100 mM phosphate, pH 7.3) freshly before the reaction or 24 hours before. If included, the persistent radical TEMPOL was added (to a 10 mM concentration). The incubated solution was either kept open to air or in a O_2-free glovebox (<6 ppm).

To this reagent (140 μL), KFWAWH peptide stock solution (5 μl, 10 mg/mL, 9.35 mM in PBS) was added and the reaction mixture was incubated on ice for 10 mins. The reaction was then initiated by addition of *t*BuOOH (1.6 μL, 72 mM in water) or was further incubated. After the desired time, an aliquot was withdrawn which was diluted with water for an LC-MS/MS analysis.

NMR investigation of NaDFES decomposition

NaDFES (3 mg/mL, 20 mM) was dissolved in d6-DMSO (550 μL) spiked with D_2O (3 μL). Another solution was prepared that additionally contained TEMPOL (0.86 mg/mL, 5 mM). Both mixtures were incubated in the NMR tubes at rt for 37 hours, during which time they were analysed periodically by NMR (relevant 1D and 2D experiments).

NMR investigation of NaDFES initiation products

Two solutions of NaDFES (3.6 mg/mL, 24 mM) were prepared in d6-DMSO (500 μL). To each of them either D_2O or (1.67 μL) or *t*BuOOH solution was added (1.67 μL, 7.2 M) and the mixtures were incubated for 1 h before the NMR analysis.

NaDFES incubation with Lys

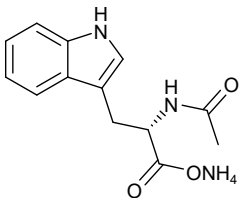
Lys·HCl (6 mg, 60 mM) and NaDFES (25.1 mg, 300 mM) were dissolved in 10x PBS buffer (0.55 mL, 100 mM, pH 7.3, prepared in D_2O). The reaction mixture was incubated for 24 hours before it was analysed comprehensively with NMR.

NaDFES incubation with melittin and proteins

NaDFES solution was prepared freshly before the reaction (up to 100 mM in 10 mM phosphate, pH 7.3). To this solution an appropriate volume of the biomolecule stock ((0.1 - 1.5) mM) was added so that the desired concentrations were achieved and the mixture was incubated for 24 h at rt. After this time, an aliquot was withdrawn which was diluted with water for an LC-MS/MS analysis.

Synthesis of TrpNAc

A glass vial was charged with Trp (25 mg, 0.12 mmol) and NaDFES (48 mg, 0.31 mmol). The solids were dissolved in DMSO (2 mL) and the reaction initiated by addition of *t*BuOOH (21 μL, 7.2 M in H_2O, 0.15 mmol). The mixture was incubated at rt for 24 h after which it was purified via automated CombiFlash chromatography on silica gel (RediSep Rf Gold 4 g Silica using gradient of aq. NH_3 (32%) in *n*-PrOH up to 15%). This allowed retrieval of white solids - the starting material (17 mg, 66%) as well as N-acetylated derivative **TrpNAc** (5 mg, 16%) as ammonium salts. Similar reaction in aqueous PBS buffer as a solvent failed to yield isolable products.



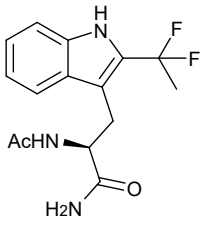
^{1}H NMR (400 MHz, DMSO-d6) δ 10.82 (d, J = 2.4 Hz, 1H, N1H), 8.13 (d, J = 7.8 Hz, 1H, HNCα), 7.52 (ddd, J = 7.9, 1.2, 0.8 Hz, 1H, H4), 7.33 (ddd, J = 8.0, 1.1, 0.8 Hz, 1H, H7), 7.13 (d, J = 2.4 Hz, 1H, H2), 7.06 (ddd, J = 8.0, 7.0, 1.2 Hz, 1H, H6), 6.98 (ddd, J = 7.9, 7.0, 1.1 Hz, 1H, H5), 4.45 (ddd, J = 8.7, 7.8, 5.1 Hz, 1H, HCα), 3.15 (dd, J = 14.5, 5.1 Hz, 1H, HCβ), 2.98 (dd, J = 14.5, 8.7 Hz, 1H, H'Cβ), 1.79 (s, 3H, $COCH_3$),

^{13}C NMR (126 MHz, DMSO-d6) δ 173.6 (COCα), 169.2 ($NCOCH_3$), 136.1 (C7a), 127.2 (C3a), 123.5 (C2), 120.9 (C6), 118.4 (C5), 118.2 (C4), 111.4 (C7), 110.0 (C3), 53.0 (Cα), 27.2 (Cβ), 22.4 ($COCH_3$),

HRMS-ESI(+): $[M+H]^+$: calcd: 247.1077, found: 247.1041.

The characterisation conforms to data from literature (WinChembase).

Synthesis of TrpCF$_2$Me



A glass vial was charged with NATA (30 mg, 0.12 mmol) and NaDFES (48 mg, 0.31 mmol). The solids were dissolved in DMSO (2 mL) and the reaction initiated by addition of tBuOOH (21 μL, 7.2 M in H_2O, 0.15 mmol). The mixture was incubated at rt for 24 h after which it was purified via automated CombiFlash chromatography on silica gel (12 g silica using gradient of MeOH in $CHCl_3$ up to 10%). This yielded **TrpCF$_2$Me** as a white solid (3.6 mg, 10%).

^{1}H NMR (400 MHz, DMSO-d6) δ 11.29 (s, 1H, N1H), 7.88 (d, J = 8.8 Hz, 1H, HNCα), 7.74 (d, J = 8.1 Hz, 1H, H4), 7.34 (d, J = 8.0 Hz, 1H, H7), 7.21 (bs, 1H, NH_2), 7.15 (dd, J = 8.0, 6.9 Hz, 1H, H6), 7.02 (dd, J = 8.1, 6.9 Hz, 1H, H5), 6.99 (s,

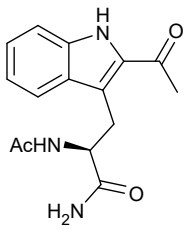
1H, NH_2), 4.46 (ddd, J = 8.8, 8.0, 6.2 Hz, 1H, HCα), 3.26 (dd, J = 14.2, 6.2 Hz, 1H, HCβ), 2.97 (dd, J = 14.2, 8.0 Hz, 1H, H'Cβ), 2.10 (t, J = 19.1 Hz, 3H, CF_2CH_3), 1.71 (s, 3H, $COCH_3$),

^{19}F NMR (376 MHz, DMSO-d6) δ −78.88 (dq, $J_{F,F}$ = 255.6 Hz, $J_{F,H}$ = 19.1 Hz, 1F), −80.03 (dq, $J_{F,F}$ = 255.6 Hz, $J_{F,H}$ = 19.1 Hz, 1F),

^{13}C NMR (126 MHz, DMSO-d6) δ 173.2 (COCα), 168.7 ($NCOCH_3$), 135.1 (C7a), 129.7 (t, J = 32.7 Hz, C2), 127.7 (C3a), 122.2 (C6), 120.7 (t, J = 235.1 Hz, CF_2), 119.5 (C4), 118.7 (C5), 111.2 (C7), 109.4 (C3), 53.5 (Cα), 26.9 (Cβ), 25.1 (t, J = 28.7 Hz, CF_2CH_3), 22.3 ($COCH_3$),

HRMS-ESI(+): $[M+Na]^+$: calcd: 332.1187, found: 332.1142.

Synthesis of TrpC=OMe



A sample of **TrpCF$_2$Me** (3 mg, 0.01 mmol) was dissolved in PBS (0.5 mL, 10 mM prepared in D_2O pH 7.2) and the solution was incubated at rt for 24 h after which, it was analysed with NMR. ^{1}H NMR indicate a complete conversion to **TrpC=OMe**, while ^{19}F NMR showed the appearance of F^- signal as the only fluorinated by-product (s, −121.4 ppm - Fig. 4.14)[238]. The compound was not isolated from this mixture. The experiment was repeated in PBS buffer that additionally contained Gly (2.3 mg, 0.03 mmol), with the same result.

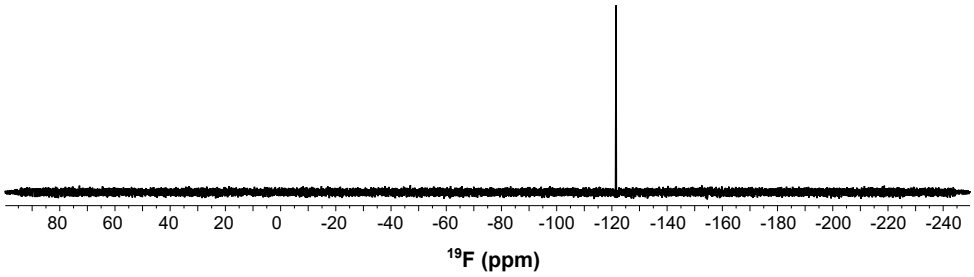


Fig. 4.14: ^{19}F NMR spectrum of the hydrolysed sample of **TrpCF$_2$Me**.

---

^{1}H NMR (500 MHz, D_2O) δ 7.78 (d, J = 8.4 Hz, 1H, H4), 7.55 (d, J = 8.3 Hz, 1H, H7), 7.46 (dd, J = 8.3, 6.9 Hz, 1H, H6), 7.25 (dd, J = 8.4, 6.9 Hz, 1H, H5), 4.60 (dd, J = 8.1, 6.4 Hz, 1H, HCα), 3.61 (dd, J = 13.9, 6.4 Hz, 1H, HCβ), 3.45 (dd, J = 13.9, 8.1 Hz, 1H, H'Cβ), 2.68 (s, 3H, $CCOCH_3$), 1.87 (s, 3H, $NCOCH_3$),

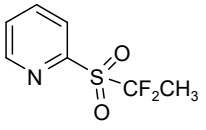
^{13}C NMR (126 MHz, D_2O) δ 196.0 ($CCOCH_3$), 176.4 (COCα), 174.4 ($NCOCH_3$), 137.0 (C7a), 132.5 (C2), 128.1 (C3a), 127.2 (C6), 121.4 (C5), 121.3 (C4), 119.2 (C3), 113.2 (C7), 54.9 (Cα), 28.3 ($COCH_3$), 27.5 (Cβ), 22.3 ($COCH_3$),

HRMS-ESI(+): $[M+Na]^+$: calcd: 310.1162, found: 310.1168,

MS/MS-ESI(+): m/z = 310.1 $[M+Na]^+ \rightarrow [M-CONH_2]^+$: calcd: 243.1134, found 243.1139.

Synthesis of PySO$_2$CF$_2$Me

A two-necked round bottom flask was charged with $PySO_2CF_2H$ (difluoromethyl 2-pyridyl sulfone, 241 mg, 1.25 mmol) and a magnetic stirring bar. The solid was dissolved in anhydrous THF (10 mL) and to this solution DMI (1,3-dimethyl-2-imidazolidinone,



1 mL) was added. The flask was cooled to −78 °C, after which MeI (120 μL, 1.88 mmol) was added. Then, LiHMDS (1.5 mL, 1.5 mmol, 1 M in THF) was added dropwise and the reaction mixture was stirred for 30 mins at −70 °C. The reaction was quenched with sat. NH_4Cl solution (20 mL) and mixture extracted with EtOAc (3 x 20 mL). The combined organic phases were washed with sat. LiCl (2 x 40 mL), dried over $MgSO_4$ and concentrated in vacuo. The residue was purified via automated CombiFlash chromatography on silica gel (RediSep Rf Gold 4 g Silica, using gradient of EtOAc in 1:1 mixture of $CHCl_3$/heptane up to 15%). This yielded **$PySO_2CF_2Me$** as a white solid (75 mg, 29%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.82 (ddd, J = 4.7, 1.7, 1.0 Hz, 1H, H3), 8.14 (dt, J = 7.8, 1.1, 1.0 Hz, 1H, H6), 8.02 (td, J = 7.8, 1.7 Hz, 1H, H5), 7.65 (ddd, J = 7.8, 4.7, 1.1 Hz, 1H, H4), 2.07 (t, J = 18.5 Hz, 3H, CH_3),

^{19}F NMR (376 MHz, $CDCl_3$) δ −96.16 (q, J = 18.5 Hz, 2F),

^{13}C NMR (126 MHz, $CDCl_3$) δ 152.1 (C2), 150.9 (C3), 138.4 (C5), 128.8 (C4), 126.4 (C6), 124.6 (t, J = 283.7 Hz, C7), 17.6 (t, J = 21.5 Hz, C8),

HRMS-ESI(+): $[M+H]^+$: found: 208.0283, calcd: 208.0238,

MS/MS-EI(+): fragmentation spectrum in the agreement with predicted species.

The characterisation conforms to data from literature[359].

Model peptide photoredox difluoroethylation

To KFWAWH solution in PBS buffer (4.6 μL, 10 mM), **$PySO_2CF_2Me$** (y μL, 1 M in DMSO), sodium ascorbate (x μL, 1.65 M in PBS buffer) and $[Ru(bpy)_3]^{2+}$ (0.7 μL, 5 mol%, 3.3 mM in PBS buffer) were added and the reaction mixture was topped up

with PBS buffer (to the total volume of 100 μL). The reaction mixture was irradiated for 15 mins after which an aliquot (2.5 μL) was diluted 10x with H_2O and analysed via LC-MS/MS. Then $NaBD_4$ (0.4 mg) was added to this solution and the reaction mixture was incubated for a further 24 h, when a final aliquot was analysed by LC-MS. In the case where 100 mM of Lys was used as the sacrificial amine, it was prepared in the PBS buffer. This PBS/Lys buffer was then subsequently used to prepare every aqueous stock solution of the other reactants.

Model protein photoredox difluoroethylation

To eqMb solution in PBS/Lys buffer (50 μL, z μM), **PySO$_2$CF$_2$Me** (y μL, 1 M in DMSO), sodium ascorbate (x μL, 1.65 M in PBS buffer) and $[Ru(bpy)_3]^{2+}$ (0.76 μL, 50 mol%, 3.3 mM in PBS buffer) were added and the reaction mixture was topped up with PBS/Lys buffer (to the total volume of 100 μL). For lower reagent excess, 10x diluted stock solutions of reagents were used. The reaction mixture was irradiated for 15 mins after which an aliquot was diluted with H_2O (to the final protein concentration of 0.03 mg/mL) and analysed via LC-MS. Reaction under denaturing conditions was performed in the buffer supplemented with urea (6 M) and after the reaction, the samples were incubated at rt for 24 h before being re-analysed.

4.7 Appendix

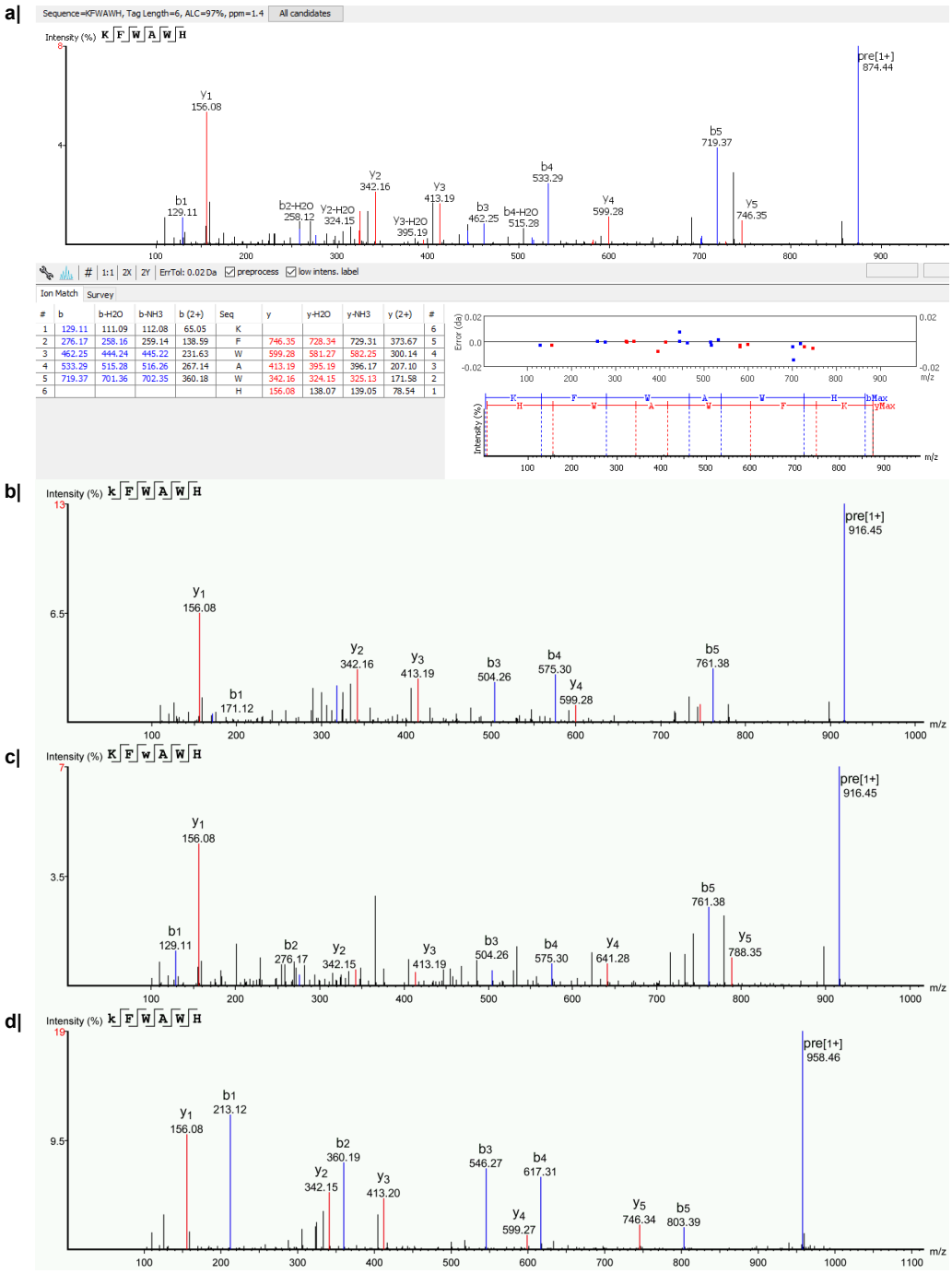
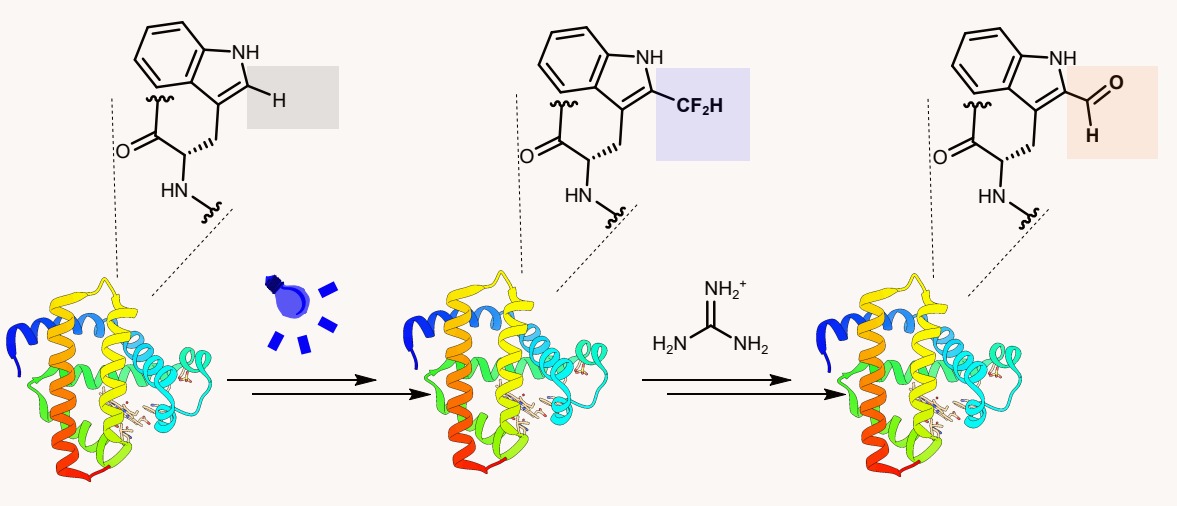


Fig. 4.15: Tandem mass spectra of different precursor ions found in the KFWAWH peptide reaction mixture with NaDFES. **a**| $m/z = 874$ ion (starting material) demonstrating original fragmentation pattern, **b**| $m/z = 916$ ion revealing Lys acetylation **c**| $m/z = 916$ ion (+42) revealing Trp acylation **d**| $m/z = 958$ ion revealing double Lys acetylation.

Chapter 5

C-Formylation of Proteins via Sequential Difluoromethylation Hydrolysis



Abstract

The results discussed in the previous chapter inspired the design of a difluoromethylation hydrolysis sequence as a novel method for introduction of an aldehyde tag into proteins. The idea is developed in Chapter 5 starting again, with the detailed selectivity and reactivity studies with small molecules and peptides. The design of the proposed transformation was verified and it was concluded that difluoromethylation occurred with a preference for tryptophan as a substrate. Investigation of peptide model allowed in-depth characterisation of possible side reactions. This allowed development of the standard conditions for photoredox difluoromethylation of a broad range of protein substrates. The hydrolytic loss of fluorine was found to be inhibited for proteins in their native fold, with the efficient transformation requiring denaturing conditions. As such, one-pot denaturative hydrolysis/labelling step was designed, and showcased in reactions with α-nucleophiles and amines, complemented with the reduction of the formed imines. In addition, characterisation of the acyl products revealed an interesting physicochemical feature of visible light fluorescence of the aldehyde tagged proteins. Together, this protocol resulted in novel bioconjugation method and a minimal size fluorescent tag introduction strategy.

5.1 Introduction

The carbonyl functional group has many distinct advantages as a chemical tag incorporated into proteins[32]. Following the discovery of the facile hydrolysis of difluoroethyl motifs attached to the native protein heteroaromatic side chains, it was hypothesised that this pathway could be used for protein C-formylation. There are a number of methods that allow the introduction of this functionality into proteins, with a prominent example of amber codon suppression enabling the installation of, for example, *para*-acetyl-phenylalanine as a handle for bioconjugation chemistry with α-nucleophiles[360,361]. However, due to their superior reactivity profile, aldehydes are typically preferred over ketones. Consequently, the genetic code expansion methods were developed for the introduction of *meta*-formyl-phenylalanine[362]. Apart from the biosynthetic methods, a carbonyl group tagging can also be achieved by chemical or enzymatic transformations. The first, can be realised by an oxidative degradation[363] of N-terminal Ser/Thr or using a commonly adopted approach of the glycan oxidation, in which certain susceptible hydroxyl groups are easily converted to aldehydes[364]. The enzymatic approach is exemplified by the action of formylglycine-generating enzymes on Cys or Gly residues in proteins with specific sequences[365,366].

While these methods have proven to be effective, they require a certain amount of the structure engineering or can be only applied in some specific cases. On the other hand, the strategy proposed here utilising difluoromethylation as a transient modification, offers a direct approach, readily applicable to a general native protein scaffold. Intuitively, a direct carbonylation using heterolytic ($2e^-$) bond-forming methods is

not feasible due to the necessity of aqueous conditions as well as presence of N- and O-nucleophiles native to proteins[59], quenching HC(O)$^{\pm}$ formyl equivalents, respectively - Fig. 5.1.

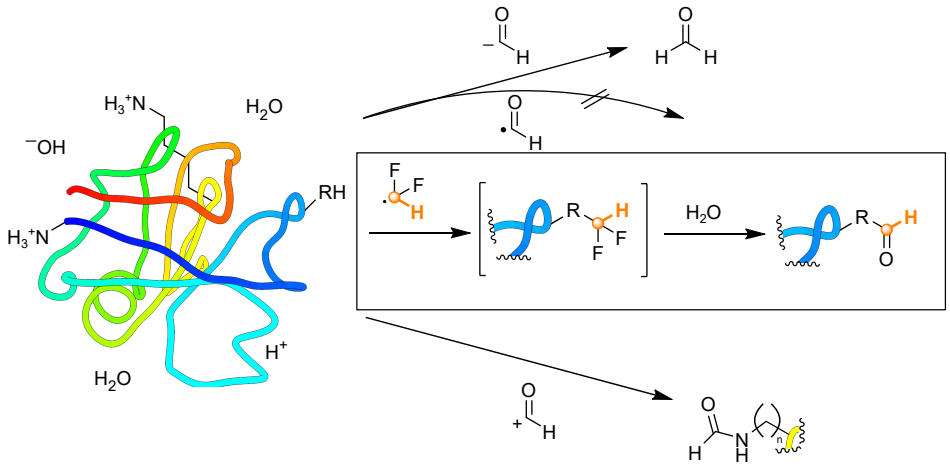


Fig. 5.1: Potential strategies for protein C-formylation with a scheme of the proposed two-step difluoromethylation/hydrolysis as a biocompatible, alternative synthetic strategy.

Similarly, a direct homolytic method cannot be applied as HC(O)• cannot be generated in a simple manner under biocompatible conditions[367]. Therefore, a two-step approach seemed to be the only feasible option and was investigated in greater detail.

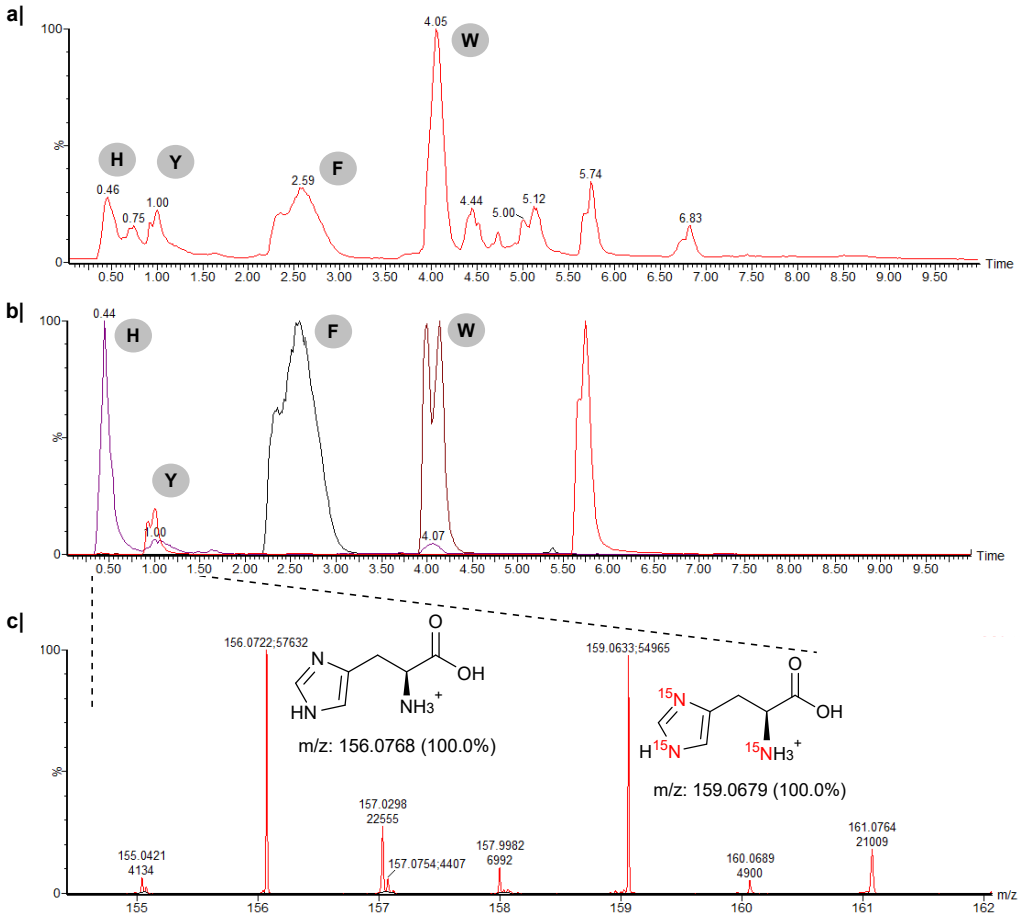
This chapter has a similar structure to Chapter 2, where reactivity and selectivity patterns are first investigated on a small molecule model. Then, the difluoromethylation of peptide substrates is presented, with a focus on the side reactions of the developed photoredox conditions. This leads to studies in protein chemistry, the verification of the difluoromethylation/hydrolysis sequence and its in-depth characterisation. Finally, potential applications of the protein products in bioconjugation and fluorescence spectroscopy are explored.

5.2 Small molecule model system

5.2.1 Establishing LC-MS assay for the AA investigation

The studies were initiated using the previously established and discussed photoredox system using pyridyl sulfone as the radical precursor, ascorbate as the reductant and $[Ru(bpy)_3]^{2+}$ as the photocatalyst (see Fig. 1.15a and Fig. 4.7). First, the reactivity of (hetero)aromatic residues was evaluated under conditions typically used for biomolecule modification: low μmol and μM range, the biomolecule substrate as the limiting species under an excess of other reactants. A quantitative, direct competition based LC-MS assay using ^{15}N labelled standards was developed to assess the engagement of different amino acids.

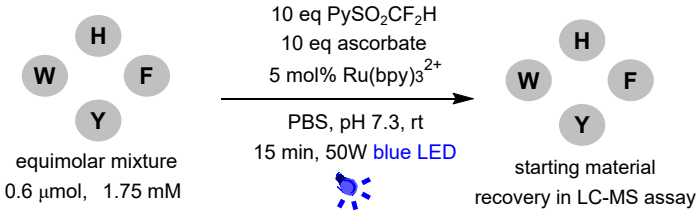
First, suitable UPLC conditions (using reversed phase column chemistry) were developed so that a full separation between amino acids of interest and other elements of the reaction mixture could be obtained - Fig. 5.2a-b. The retention time (RT) of the individual amino acids was established by a co-injection of ^{14}N and ^{15}N standard samples. No other signals for the m/z of interest (Fig. 5.2 Table, within $\sim$100 ppm) were observed in the combined chromatographic regions, as verified by the blank injections and examination of the extracted ion chromatograms (XIC) for each of the species. The relative quantification was achieved by the comparison of natural and labelled AAs signals (the latter added just before the analysis in a known quantity) - Fig. 5.2c.



Isotope	m/z Calc.	m/z Observed	AA	RT region [min]
^{14}N	205.0972	205.0915	Trp	3.90 - 4.50
^{15}N	207.0912	207.0848		
^{14}N	166.0863	166.0801	Phe	2.00 - 3.50
^{15}N	168.0866	168.0810		
^{14}N	182.0812	182.0754	Tyr	0.90 - 1.50
^{15}N	184.0816	184.0758		
^{14}N	156.0768	156.0722	His	0.25 - 0.85
^{15}N	159.0679	159.0633		

Fig. 5.2: **a**| TIC chromatogram of the AA reaction mixture (before light irradiation) with marked RTs, **b**| XIC of relevant AA masses as per table below, **c**| Combined mass spectrum of RT corresponding to the peak of His. Table: description of different AAs signals chosen for the analysis.

Having established the assay, an equimolar mixture of the putative reactive amino acids (Trp, His, Tyr and Phe) was tested under the proposed photoredox conditions. The results suggested that only Trp was substantially consumed (30%) in such conditions, while much lower reactivity was observed for His - Tab. 5.1.



Tab. 5.1: LC-MS quantification of competitive AA reactivity under photoredox difluoromethylation conditions. The conversion was based on a relative signal intensity of the starting material with a ^{15}N reference correction.

		Average area				
Isotope	**AA**	**hν−**	**hν+**	**Ratio**	**Remain**	**Conversion**
Experiment 1						
^{14}N	Trp	92259	79876	0.866	70%	**30%**
^{15}N		235424	289272	1.229		
^{14}N	Phe	322942	359890	1.114	99%	1%
^{15}N		82491	93285	1.131		
^{14}N	Tyr	128793	167624	1.301	101%	-1%
^{15}N		8330	10759	1.292		
^{14}N	His	71111	81471	1.146	98%	2%
^{15}N		41348	48401	1.171		
Experiment 2						
^{14}N	Trp	92629	61253	0.661	69%	**31%**
^{15}N		405819	386804	0.953		
^{14}N	Phe	378596	328496	0.868	100%	0%
^{15}N		132013	114972	0.871		
^{14}N	Tyr	182500	171377	0.939	98%	2%
^{15}N		13435	12857	0.957		
^{14}N	His	57877	50130	0.866	96%	4%
^{15}N		55411	49866	0.900		

On the other hand, both Phe and Tyr were essentially unaffected. To further verify this result, the experiment was repeated using a higher reagent excess and a different analytical method, in which the UV absorption was used for the quantification of the amount of starting material consumed. The conclusions were confirmed, as ~29% of Trp reacted under these conditions, suggesting a full engagement of the catalyst during the irradiation or a maximum use of the photon flux - Fig. 5.3. Taken together, it was predicted that protein difluoromethylation may be possible with a high degree of Trp chemoselectivity.

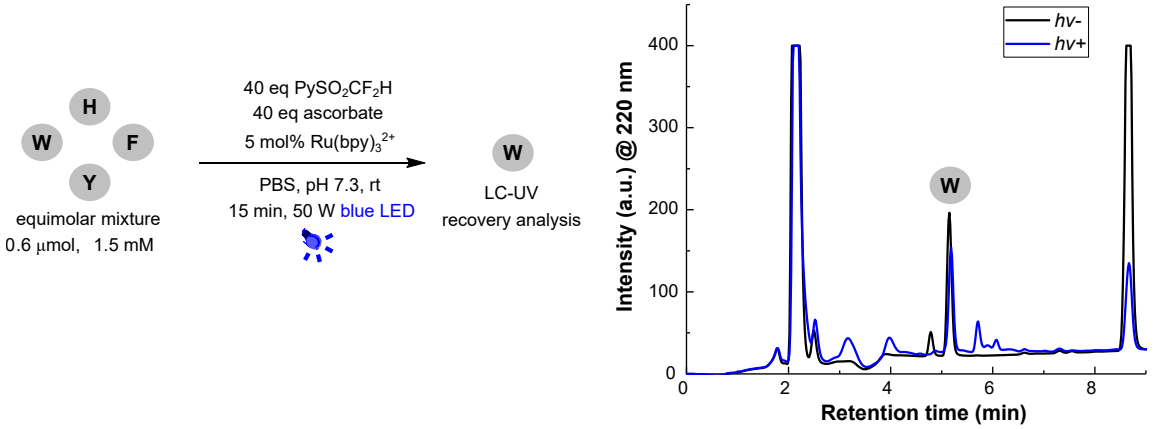


Fig. 5.3: LC-UV quantification of competitive Trp reactivity under photoredox difluoromethylation conditions. Position of Trp peak at RT = 5.1 min was established by injection of the reference. The conversion was based on the recovery of the starting material by integration of the area under the peak (fitted by a Gaussian function in Origin v9.0).

5.2.2 Verifying reactivity

A similar LC-MS assay under competitive conditions was used for an evaluation of products formed in the reaction. A differential mass spectrum analysis (obtained via subtraction of spectra for non-irradiated and no AA reaction mixtures from the combined mass spectrum) revealed the appearance of several new species, most of

which could be identified through the comparison with expected monoisotopic masses. Interestingly, products of the radical addition were not the most abundant and the hydrolysed acyl derivatives (for Trp and His) were not observed at all - Fig. 5.4a.

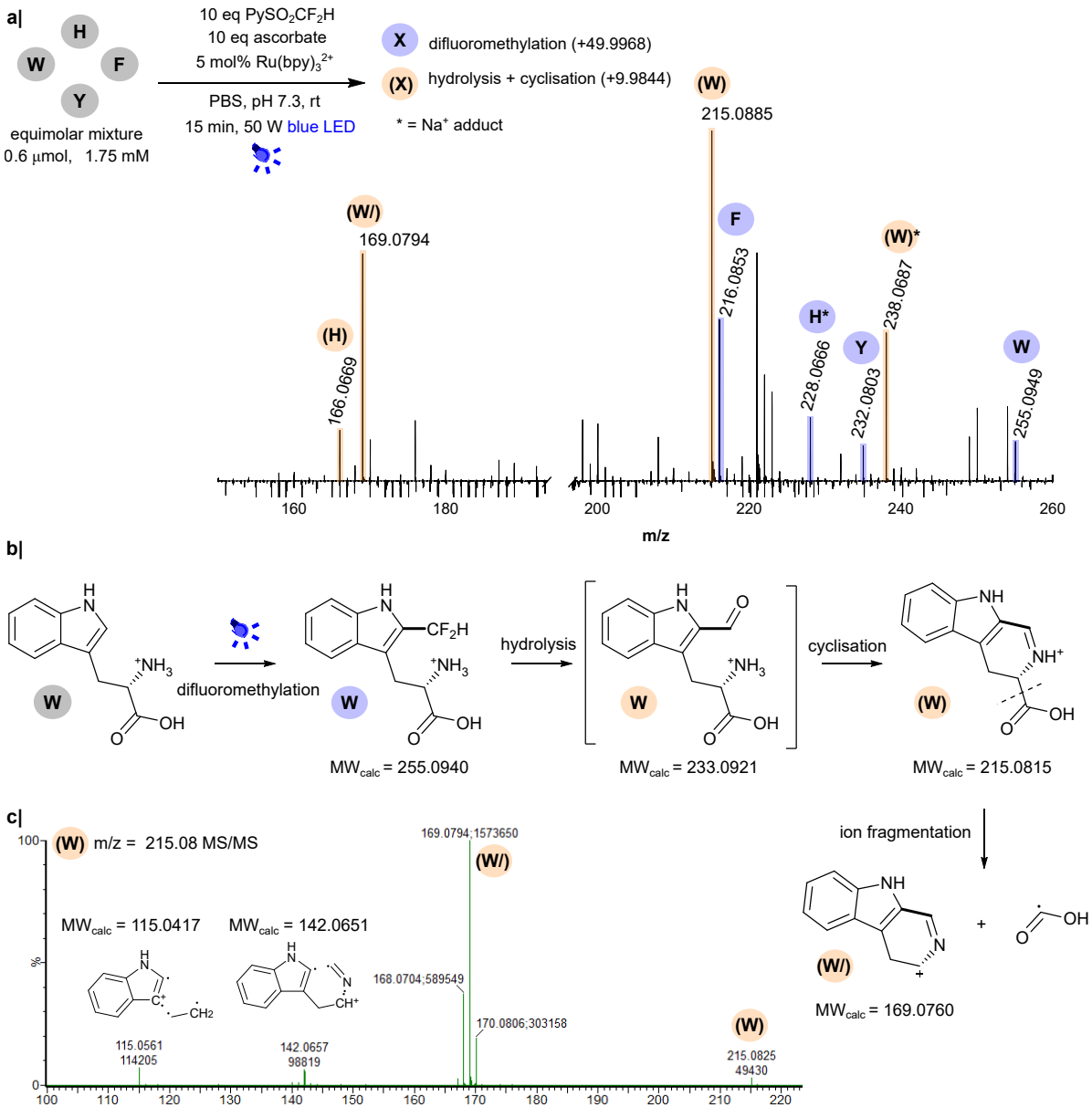


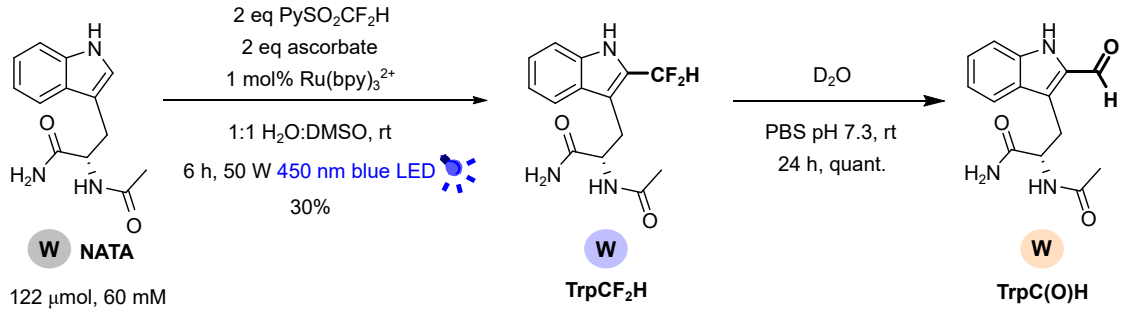
Fig. 5.4: Difluoromethylation of AA mixture. **a**| A differential mass spectrum with peaks annotated with potential products based on their expected masses. **b**| Proposed scheme for the cyclic aldimine formation alongside its MS fragmentation product (labelled as (W/)). **c**| Tandem mass fragmentation spectrum (under collision induced dissociation (CID) conditions with collision energy of 10 V) with a proposed fragments.

Instead, intensive peaks with masses corresponding to the presumably cyclised imine products were observed. For Trp, the most likely origin of such species was via modification at the C2 position, where the resulting aldehyde would be well poised for internal imine formation - Fig. 5.4b. This hypothesis was confirmed by the analysis of the fragmentation spectra - Fig. 5.4c (similarly for His, not shown) and the comparison with similar transformations reported in the literature[368].

Next, the formal confirmation of the difluoromethylation/hydrolysis regioselectivity and reactivity for Trp was achieved by the investigation of a small molecule model. In order to avoid the complication of the internal imine formation, the N-protected amino acid derivative - NATA (*N*-acetyltryptophanamide) was used as the substrate, which would allow the direct isolation of the aldehyde derivative. Unfortunately, in fully aqueous conditions the substrate was only partially soluble and irradiation resulted in the formation of a complex mixture, for which the products could not be isolated (data not shown).

On the other hand, using semi-aqueous conditions with DMSO as a co-solvent inhibited the hydrolysis, and allowed isolation of the C2-difluoromethylated species as the main product (according to the LC-UV analysis of the reaction mixture at 280 nm characteristic for indole derivatives, not shown). The reaction under these conditions also led to the generation of low amounts of several products (different regioisomers and oxidised derivatives as characterised by LC-MS) that were not separated (<20%), alongside a substantial amount of the unreacted starting material. Pleasingly, isolated **TrpCF$_2$H** product was cleanly hydrolysed to the formyl product **TrpC(O)H** upon a prolonged incubation in a D_2O buffer, without deuteration, conclusively confirming

the predicted reactivity - Sch. 5.1.



Sch. 5.1: Difluoromethylation of NATA and subsequent hydrolysis in a deuterated buffer.

5.3 Peptide model

5.3.1 Mechanistic investigations on a model peptide

Next, the reaction was studied on the peptide model to further understand the characteristics and to confirm the observed selectivity. H_2N-KFWAWH-COOH, used previously for the difluoroethylation studies, was chosen as the substrate as it contained the putative reactive residues of interest. Initially, it was verified that all reaction components were necessary for the observed difluoromethylation reactivity. Under such conditions, the expected peaks for the difluoromethylated and hydrolysed products were detected by LC-MS - Fig. 5.5a. Expectedly, no products were observed in the absence of light or photocatalyst (data not shown). The presence of ascorbate was instrumental not only in the radical generation but also as a potent antioxidant, as evidenced by the appearance of the oxidation pattern in the mass spectrum in its absence - Fig. 5.5b. Used chromatographic conditions allowed for the separation of different difluoromethylated products, revealing that the majority of the modification was localised on the either of the two Trp residues, with 8x excess over the residual

His products signals - Fig. 5.5c. The modification sites were determined by analysis of the tandem mass fragmentation spectra - Appendix (Fig. 5.33 - Fig. 5.35).

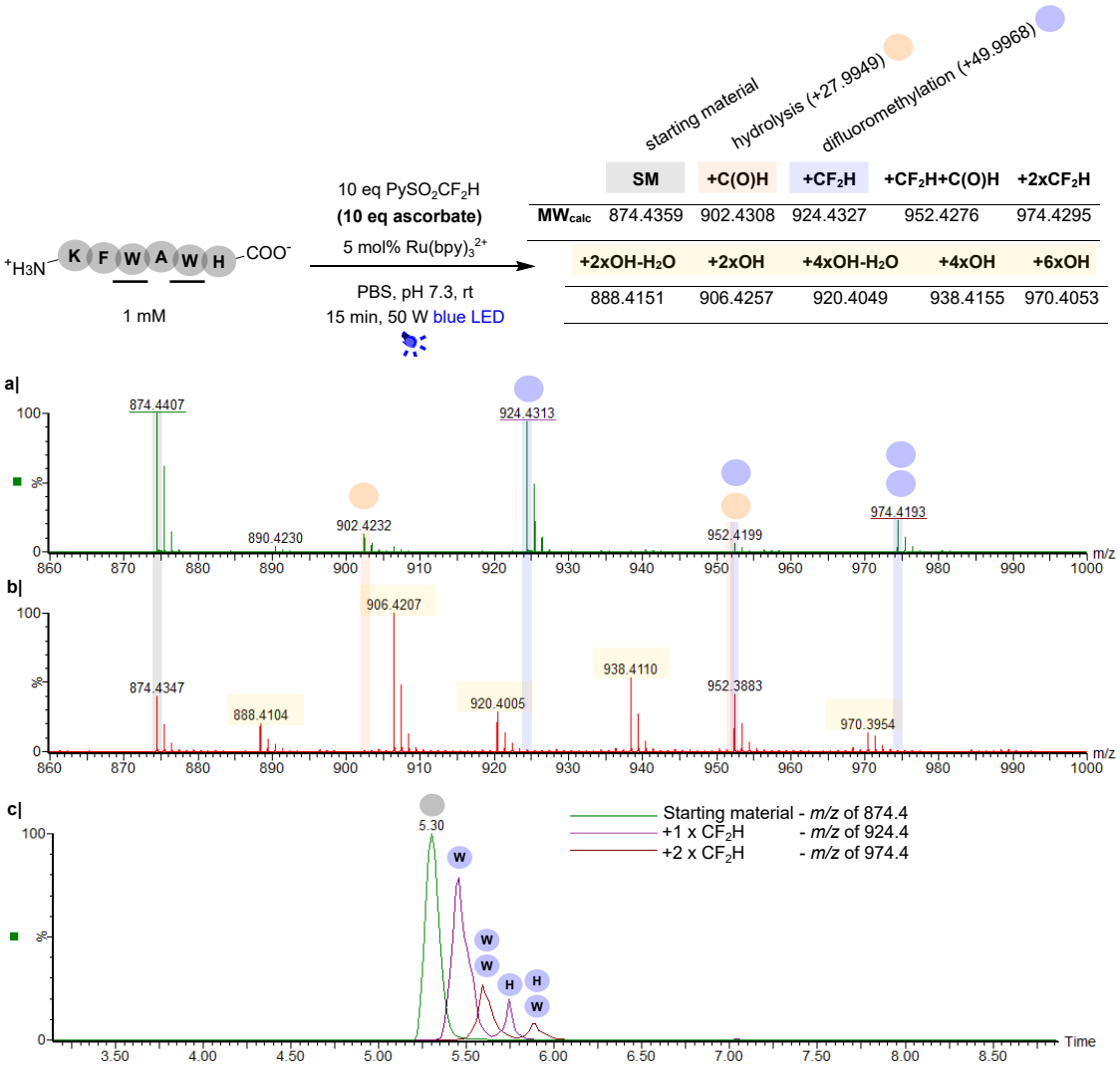


Fig. 5.5: Photoredox difluoromethylation of the model peptide. Combined mass spectra of the reaction mixture **a**| in the presence and **b**| in the absence of ascorbate as a reductant. A putative assignment of products was based on the molecular formulae change as presented in the table. **c**| XIC of starting material as well as single and double difluoromethylated products. This spectrum was obtained for a reaction in which a higher excess of reagents was used (25 eq of $PySO_2CF_2H$ and 25 eq of ascorbate).

To confirm the radical mechanism, a persistent radical TEMPOL was added to the reaction mixture - under such conditions, only minor amounts of products were ob-

served - Fig. 5.6a. Instead, a prominent TEMPOL–CF_2H adduct could be detected, as verified by the HR-MS and analysis of the fragmentation pattern - Fig. 5.6b.

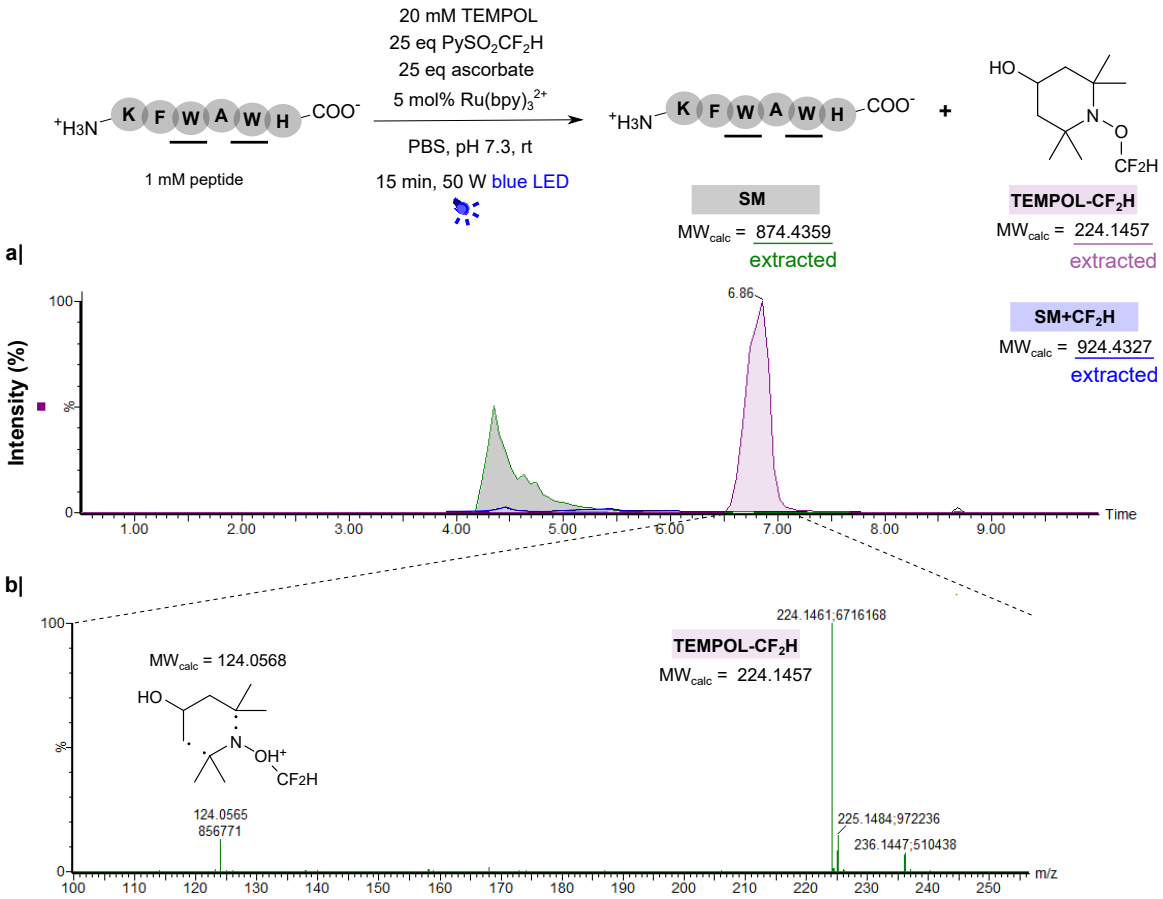


Fig. 5.6: Inhibition of the model peptide difluoromethylation by TEMPOL. **a**| XIC showing ions corresponding to the peptide starting material, difluoromethylated product as well as the TEMPOL–CF_2H adduct. Note different than usual chromatographic conditions (reversed phase phenylated stationary phase). **b**| A tandem mass spectrum of the adduct (m/z = 224.1) confirmed the compound's identity based on the proposed interpretation of the fragmentation pattern.

To test if ascorbate was an optimal reactant for this transformation, other antioxidants and reductants were screened, which revealed the superior role of ascorbate - Fig. 5.7.

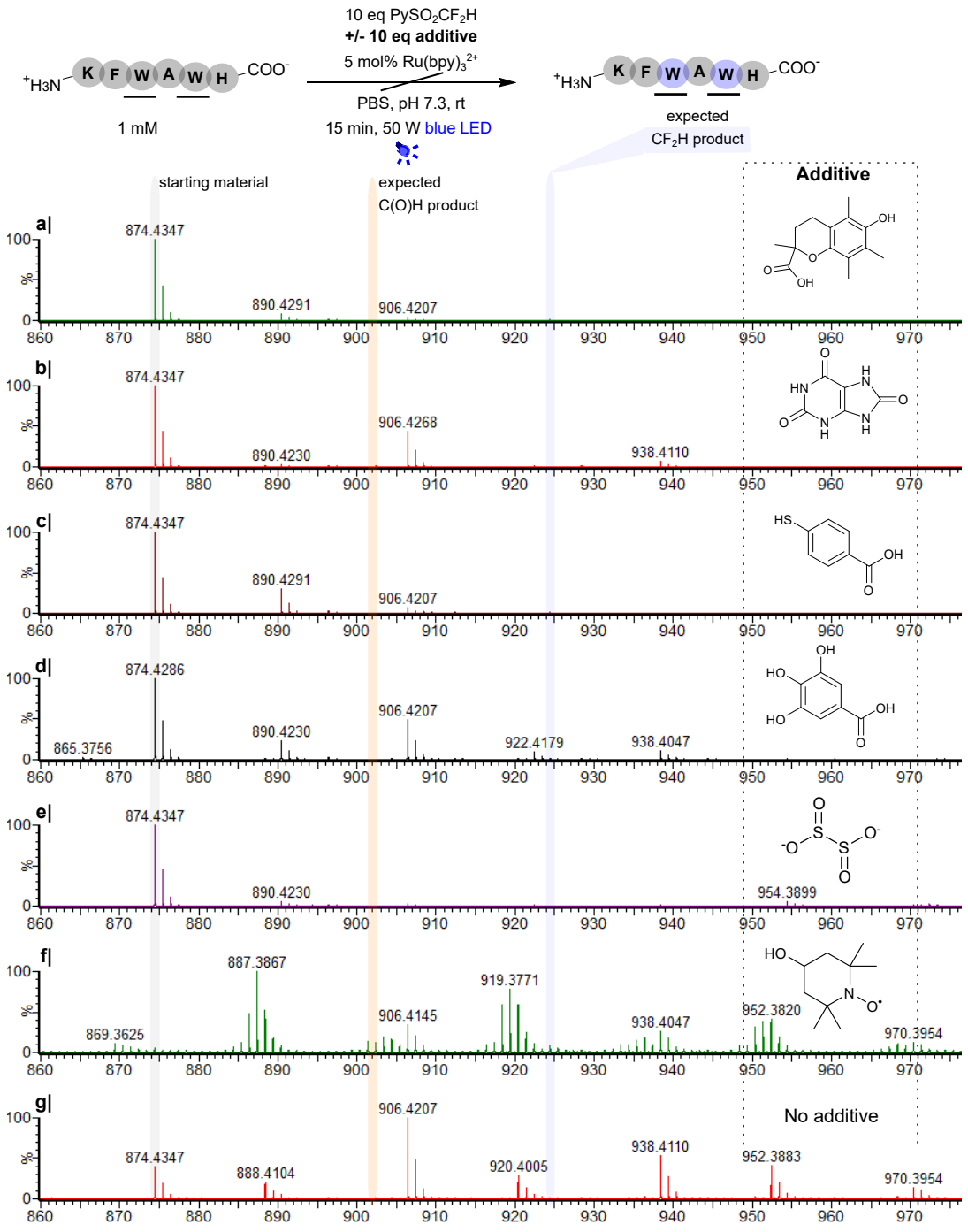


Fig. 5.7: LC-MS screening of additives for photoredox difluoromethylation of the model peptide. **a-g|** Combined mass spectra of the reaction mixtures with various additives (structures shown) a) trolox, b) uric acid c) 4-mercaptobenzoic acid, d) gallic acid, e) sodium dithionite, f) TEMPOL, g) no additives.

None of the tested redox active agents were capable of efficiently inducing the radical

formation, but most of them suppressed the observed oxidation.

Having verified the basic difluoromethylation reactivity, further experiments were aimed at studying the hydrolysis reaction. To evaluate the kinetics of this process and a possible influence of acid/base catalysis, a sample of difluoromethylated peptide was incubated in reaction mixtures at different pH values and periodically analysed by LC-MS - Fig. 5.8.

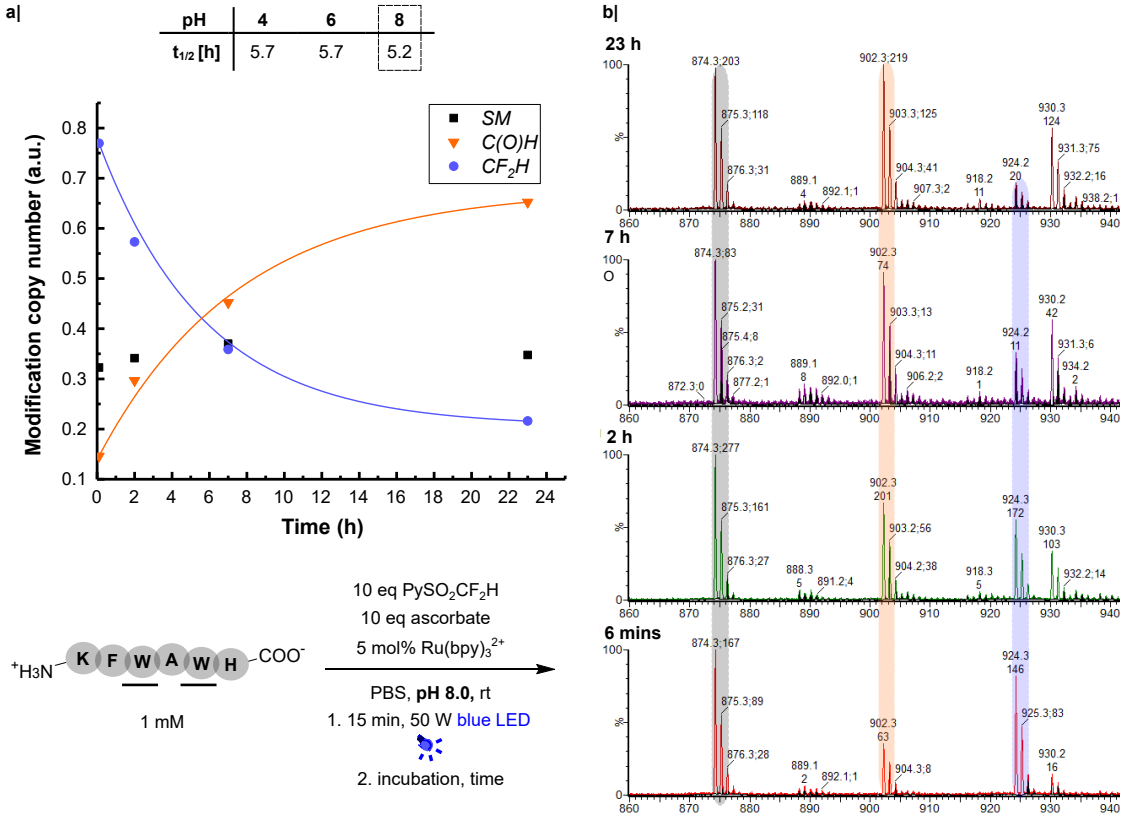


Fig. 5.8: Investigation of the difluoromethyl modification hydrolysis for a model peptide incubated in the reaction mixture. **a**| Time-course quantification of formylated and difluoromethylated products change revealed an approximate first order rate, from which the half-life at different pH values was estimated. $[M+2H]^{2+}$ integrated peaks were used to calculate the relative modification ratios (due to the higher peak intensity), here shown for pH 8. **b**| Combined mass spectra of the reaction mixture (pH 8) periodically analysed with LC-MS - analysed modification peaks colour coded.

It was found that the hydrolysis was not significantly different in the range of tested conditions and that it could be approximated by a first order kinetics, giving a fitted half-life of <6 h.

5.3.2 Peptide model for probing selectivity and N-reactivity

Having confirmed the facile hydrolysis of the CF_2H tag in the peptide model, the residue distribution of the acyl products was investigated - Fig. 5.9. Shortly after the difluoromethylation reaction, two main populations of acylated products could be observed and MS/MS analysis established that they corresponded to the two main products with modification on Trp or Lys - Appendix (Fig. 5.36 and Fig. 5.37).

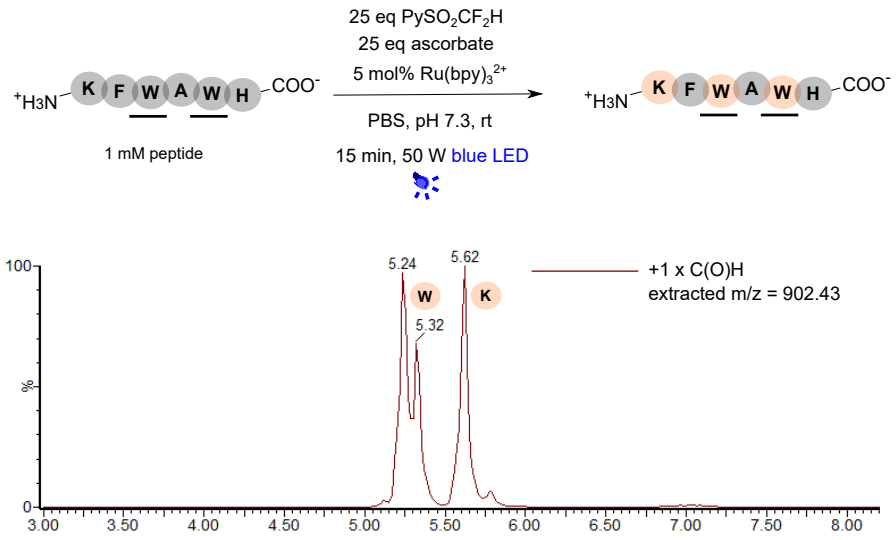
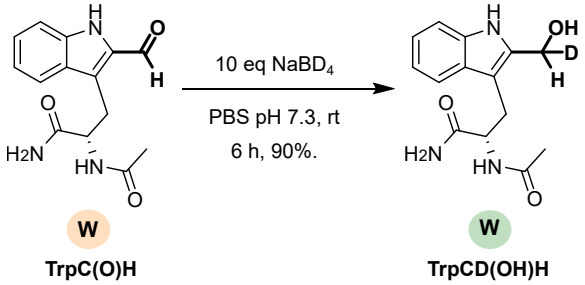


Fig. 5.9: Analysis of the acylated products for the model peptide difluoromethylation. XIC of ions corresponding to the single formyl modification revealed two main populations of different products that could be assigned by MS/MS analysis (the observed doubling of the Trp peak was due to the modification of either one of these residues present).

This result was not unexpected, as a certain level of non-specific N-reactivity was predicted based on the previous of studies on difluoroethylation using a similar precursor. Those studies revealed the formation of an unstable sulfinate under the reaction condi-

tions, which ultimately formed a reactive agent - acyl fluoride (Chapter 4). However, in this case, the corresponding compound would be formyl fluoride which is known to be hydrolytically unstable and volatile[369] - accordingly, only low levels of formylation were detected (the intensity of Lys formylated product constituted <1% of the total peptide species ion count).

To confirm this result, the borodeuteride reduction assay (developed in Chapter 4.4.1) was applied. First, the reactivity was tested on a small molecule model, where treatment of the **TrpC(O)H** with 10 eq of the reagent resulted in 90% conversion to the expected product - Sch. 5.2.



Sch. 5.2: Reduction of **TrpC(O)H** with sodium borodeuteride.

Next, attention was turned to the model peptide. To achieve full hydrolysis of the difluoromethyl groups, the reaction mixture was incubated for 24 h before exhaustive reduction with $NaBD_4$. LC-MS monitoring of the reaction at each stage, revealed the expected patterns of reactivity, while MS/MS analyses confirmed the Trp selectivity of the whole sequence - Fig. 5.10. This experiment suggested that the same reaction sequence targeting Trp residues may be possible in proteins. To gain further insights into reactivity when Trp residues are absent (or heavily outnumbered by other reactive residues), an analogous study was performed for the Trp-free version of the model

peptide. Consequently, a H_2N-KFMAYH-COOH peptide was synthesised, where both Trp residues were replaced with other potentially reactive residues: Tyr which could be intrinsic radical acceptor, as well as Met which are known to be sensitive to oxidative stress, possibly arising from the photoredox catalysis in the presence of oxygen.

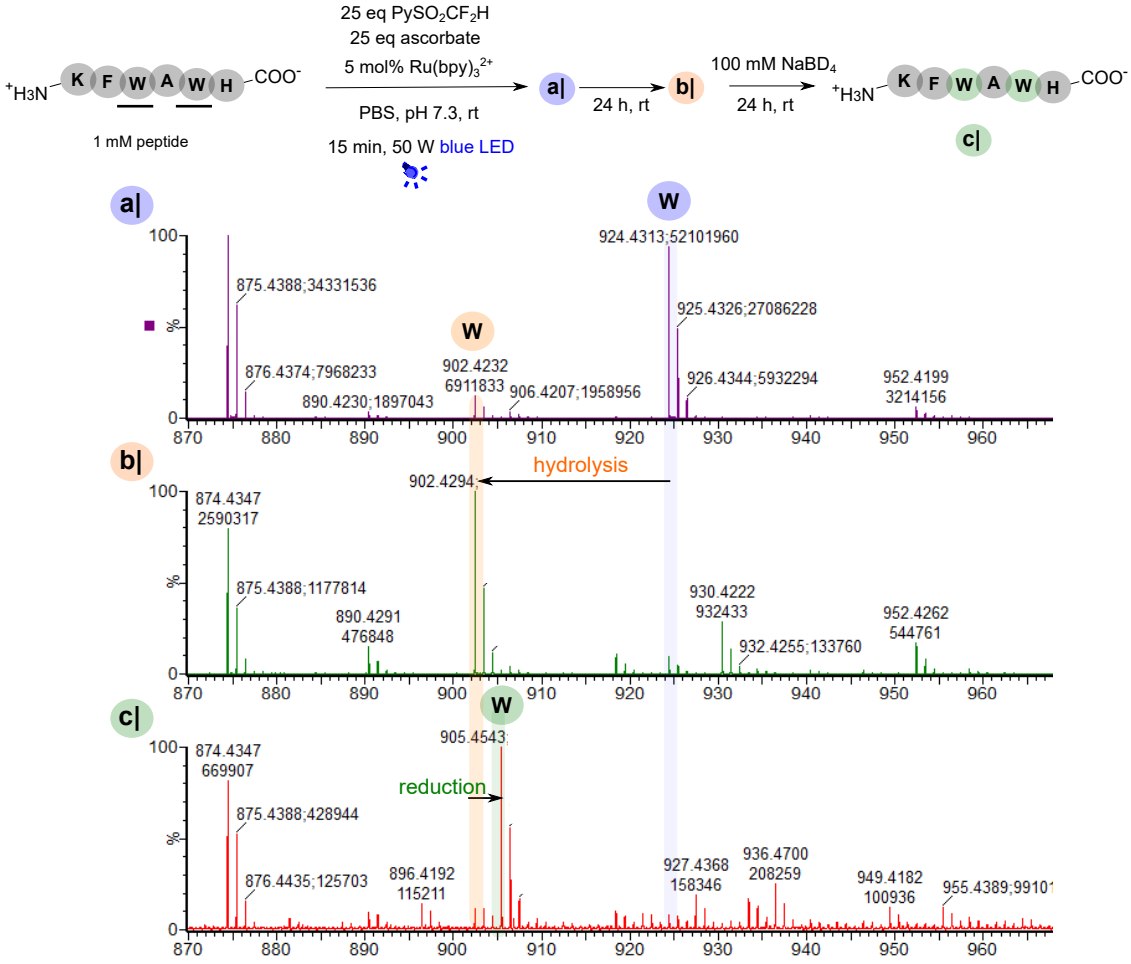


Fig. 5.10: LC-MS analysis of the difluoromethylation/hydrolysis/borodeuteride reduction sequence for the model peptide. **a-c|** Combined mass spectra of the reaction mixture after each respective stage: (a) difluoromethylation, (b) hydrolysis (c) reduction. The main, colour coded products peaks were traced according to the predicted reactivity. Note that higher mass peaks (> 925 Da) originate from the double modification but follow the same pattern of reactivity (due to the presence of 2 Trp residues).

The reaction sequence was repeated for the KFMAYH peptide and LC-MS analysis revealed that difluoromethylated products were formed in minimal quantities, identified on both His and Phe residues - Fig. 5.11a and Appendix (Fig. 5.38 - Fig. 5.40). Interestingly, their intensity throughout the sequence did not change, suggesting that hydrolysis of the tag was specific to Trp residues - Fig. 5.11b.

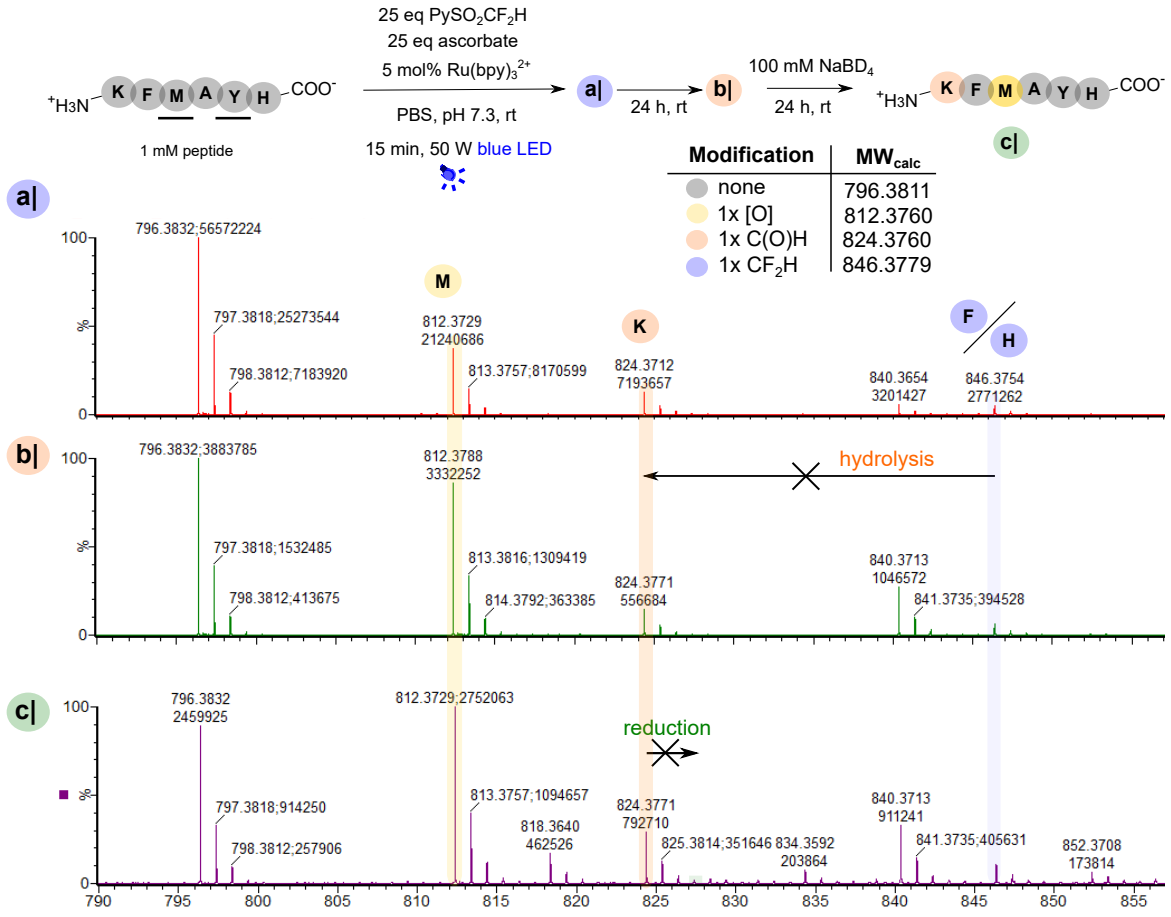


Fig. 5.11: LC-MS analysis of the difluoromethylation/hydrolysis/borodeuteride reduction sequence for the Trp-free model peptide **a-c**| Combined mass spectra of the peptide are shown after each respective stage: (a) difluoromethylation, (b) hydrolysis (c) reduction. Note that the peak at $m/z = 840.3$ was elucidated to originate from Met3 oxidation & Lys1 formylation.

Similarly, only a low intensity formylation product could be observed, with MS/MS

analysis confirming its localisation on the N-terminal Lys. As expected, the treatment with borodeuteride did not introduce any changes to the composition of the reaction mixture, with the persistent formylation product peak still present and no additional $\Delta m = +3$ Da peaks emerging - Fig. 5.11c. However, a prominent peak with a mass corresponding to the oxidation product could be observed - the intensity of which was increased over time. MS/MS analysis confirmed its predicted site within the Met residue, suggesting that solvent exposed Met residues in proteins are likely to be susceptible to oxidation under the proposed protocol.

Before moving onto proteins, difluoromethylation specificity was confirmed for another Trp/Trp-free peptide pair - H_2N-YPWF-$CONH_2$ and H_2N-YPWF-$CONH_2$ - Fig. 5.12.

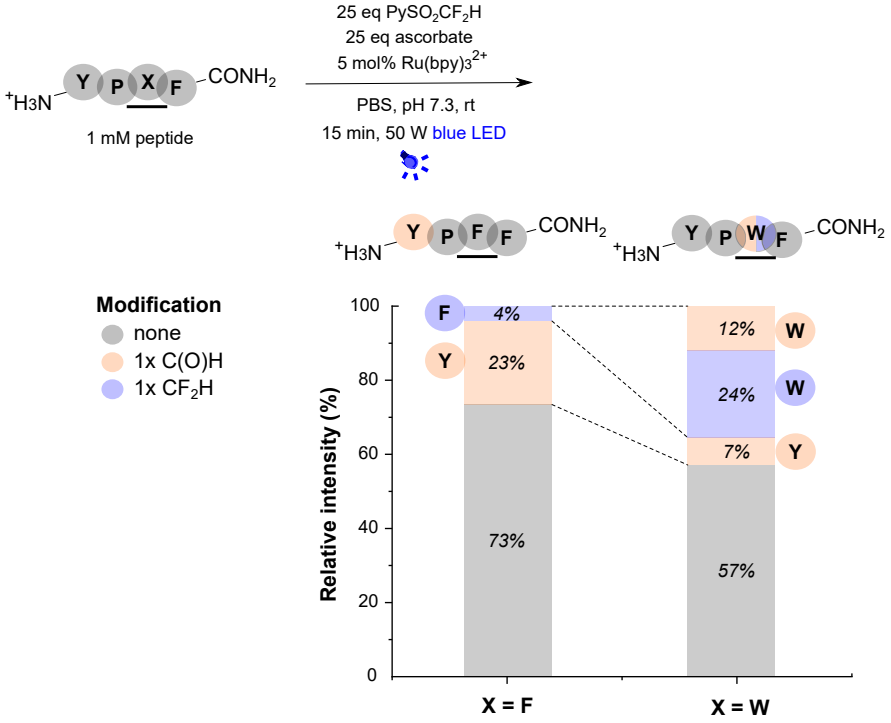


Fig. 5.12: A comparison of normalised product distribution for difluoromethylated YPFF and YPWF peptides.

LC-MS analyses of the reaction mixtures confirmed previous observations. The Trp containing YPWF substrate yielded both the prevalent CF_2H and C(O)H modifications of this residue, while the Trp-free YPFF showed very low background difluoromethylation of Phe, as well as formylation of the N-terminus (Tyr). Overall, the studies carried out on the model peptides and small molecule substrates, highlighted the specificity of the difluoromethylation/hydrolysis sequence for Trp residues and confirmed much lower reactivity of other aromatic residues. The results of extensive LC-MS analyses and the reductive assay supported these findings but also revealed the possible of N-formylation as a background side reaction; however, it was hypothesised that the use of a sacrificial amine in the buffer, as found for the difluoroethylation, would suffice to curtail it (see Fig. 4.10). The evaluation of this hypothesis was performed for proteins.

5.4 Protein chemistry

5.4.1 Difluoromethylation

To verify the suitability of the difluoromethylation/hydrolysis sequence for proteins, initial investigations were performed using myoglobin, the previously studied model protein. First, a simple difluoromethylation experiment was performed, and LC-MS analysis of the reaction mixture revealed the formation of expected peaks, corresponding to formylated and difluoromethylated protein - Fig. 5.13a. Tryptic digest and peptide mapping established that the majority of the formylation was localised within the N-terminus and Lys residues, as opposed to the Trp residues, suggesting a significant undesired N-reactivity - Appendix (Fig. 5.41). To restrain this reactivity,

the addition of different competitive agents was studied. As concluded previously, the inclusion of 100 mM Lys as a sacrificial amine, was enough to significantly reduce this reaction - Fig. 5.13a.

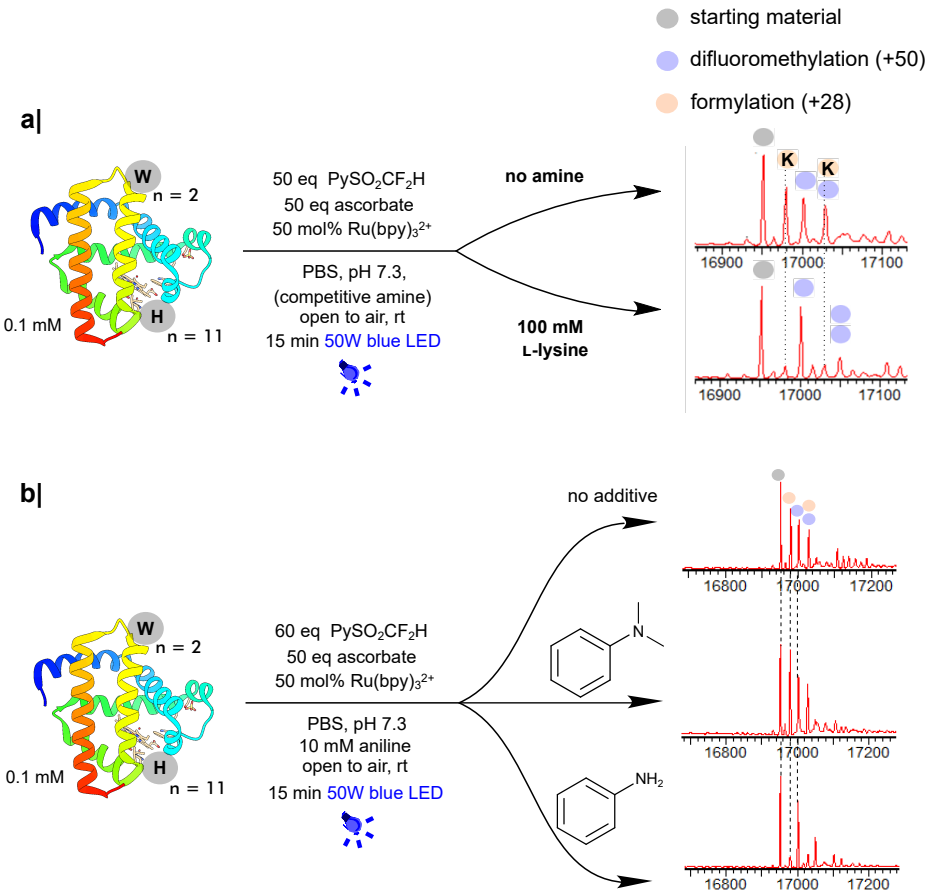


Fig. 5.13: Comparison of competitive amines influence on difluoromethylation of the model protein myoglobin. **a**| LC-MS analysis revealed that the addition of Lys inhibited the formation of the formylated protein products. **b**| The presence of a nucleophilic aniline resulted in the absence of formylated protein peaks in the mass spectra.

Similarly, difluoromethylation as a near exclusive modification was also observed in the presence of aromatic amine, while the non-nucleophilic, dimethyl aniline did not have this protective effect - Fig. 5.13b. Due to the solubility reasons, (aniline stock needed to be prepared using DMSO) and possible interference in the difluoromethyla-

tion reaction (aromatic ring participation in radical trapping), Lys was chosen as the sacrificial amine of choice. A short optimisation of the reagent excess led to the development of standard conditions, under which conversions of >60% were achieved for several proteins with different numbers of Trp residues and structure characteristics - Fig. 5.14a.

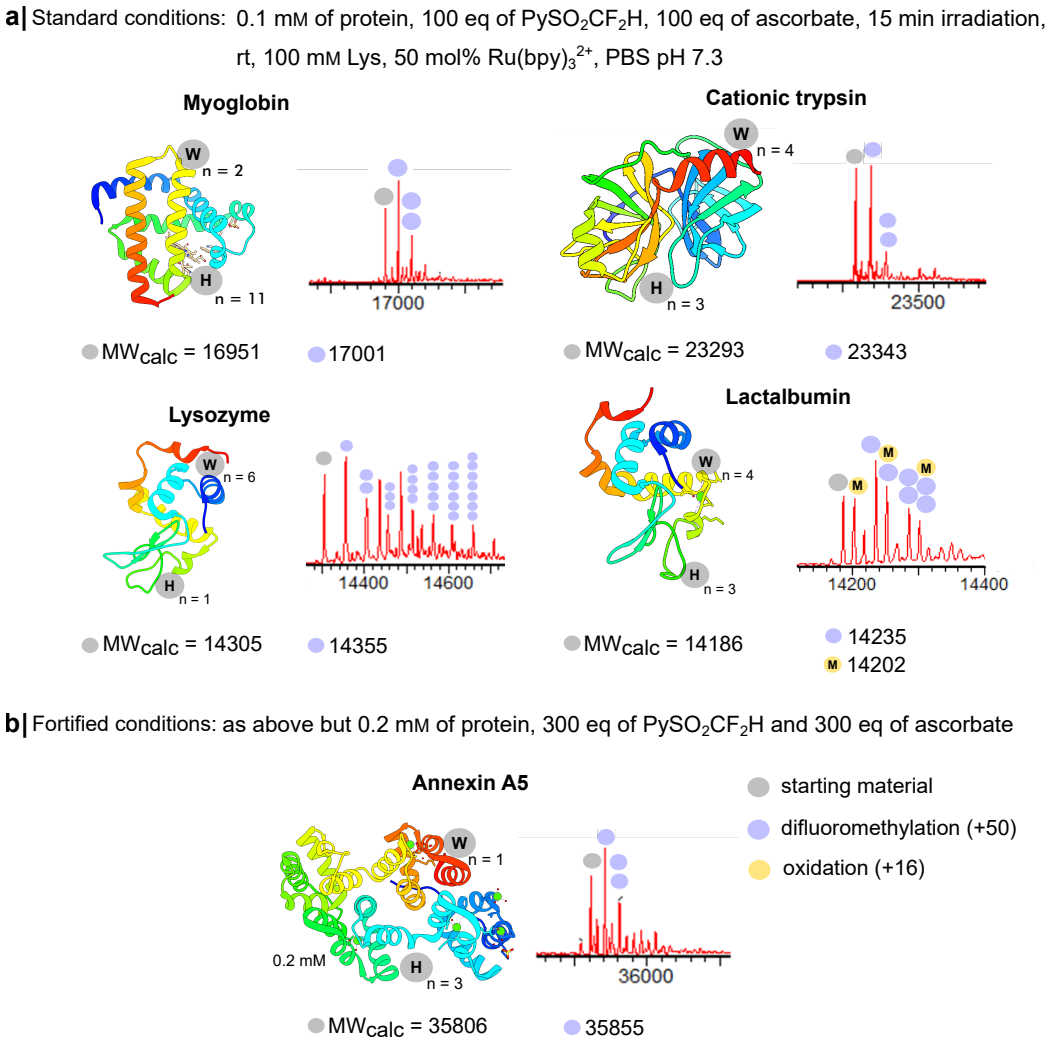


Fig. 5.14: LC-MS analyses of difluoromethylation of different proteins. **a**| Under standard conditions, a range of proteins yielded expected products as by the appearance of peaks of expected mass. In the case of lactalbumin, an additional set of peaks corresponding to oxidation products (marked with yellow) was observed, **b**| For a protein with a single Trp residue, fortified conditions were required to achieve higher conversion to products.

As such, predominantly α-folded myoglobin; two mixed α/β domains proteins: lysozyme[370] (Ly) and lactalbumin[371], as well as predominantly β-structured trypsin[372], were all efficiently difluoromethylated, with Ly displaying a high copy number of CF_2H modification, in accordance with the highest abundance of Trp residues (6). Not unexpectedly for lactalbumin, which has one solvent accessible Met[371], oxidation products were also observed. For a protein with a single Trp (annexin V - A5), a rather low conversion was obtained under such conditions and a higher excess of the reagents was needed to increase it, at the cost of the appearance of a small population of double (hence most likely non-Trp localised) difluoromethylated products - Fig. 5.14b.

The standard set of conditions developed here operated at neutral pH, which are typically preferred for protein modification. However, this was hypothesised to contribute to the Trp-selectivity. Indeed, it has been reported that similar fluoroalkylations of peptides performed at lower pH exhibited strong His-selectivity, presumably driven by the Minisci-type reactivity of the protonated side chains[78]. Supporting this hypothesis, enhanced conversions were observed when myoglobin was difluoromethylated in buffers at lower pH values - Fig. 5.15. Under the standard conditions proposed here, species with an exceeding copy-number (>2) of difluoromethyl tags emerged in a more acidic reaction mixtures, implying that pH was indeed partially responsible for the observed selectivity and the suppression of the nucleophilic nature of the CF_2H radical.

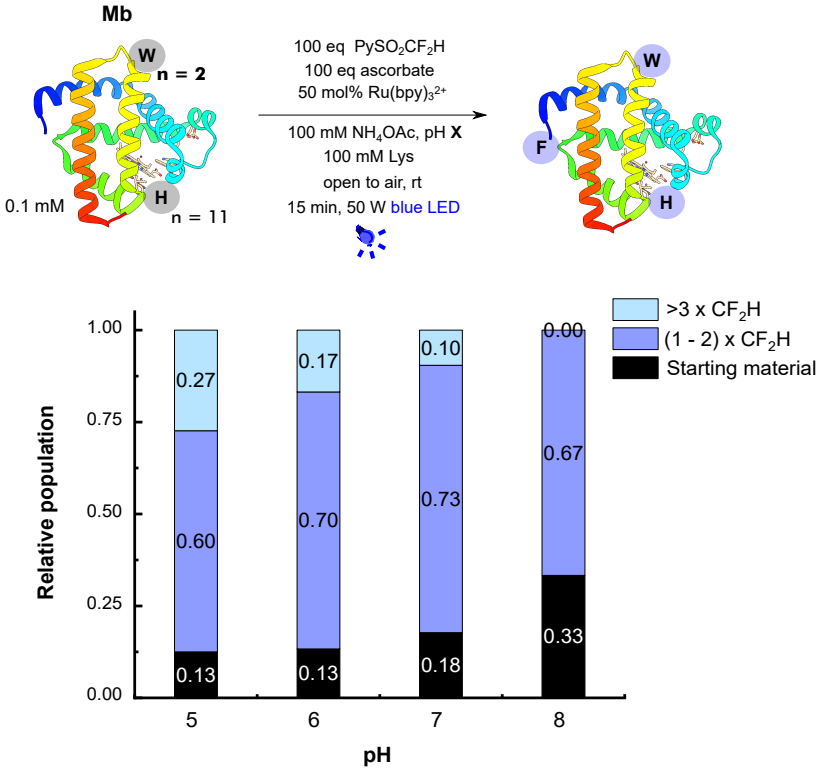


Fig. 5.15: Difluoromethylation of myoglobin in buffers of different pH. Relative populations of species with expected (1-2) and excess (≥3) content of difluoromethyl groups were quantified by LC-MS analysis.

Investigation of reactions with variable amounts of $[Ru(bpy)_3]^{2+}$ revealed that truly catalytic quantities were sufficient to obtain efficient reaction - Fig. 5.16, which is not a typical situation in protein chemistry[373]. However, across a range of protein substrates, better conversions were obtained using 50% mol of the photocatalyst. This substoichiometric amount was sufficient to counteract the presence of different chromophores competing for light (as myoglobin absorbs in a similar spectral region[248]) and was chosen to ensure the robustness of the protocol.

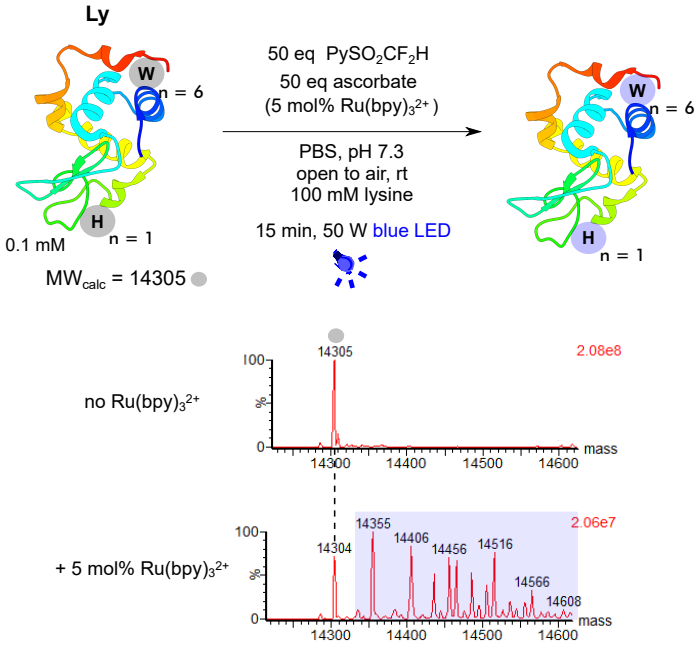


Fig. 5.16: LC-MS analysis of lysozyme difluoromethylation with and without substoichiometric amounts of photocatalyst. In its absence only the starting material was observed, while addition of as low as 5 mol% resulted in the expected difluoromethylation ladder.

To investigate the influence of sterics and secondary structure on the outcome of the reaction, the procedure was carried out under denaturing conditions. In high concentration of urea, lower CF_2H copy number was observed for difluoromethylation of myoglobin - Fig. 5.17a. Less perturbing conditions could be achieved for annexin 5, for which the presence of Ca^{2+} (see Fig. 2.21), specifically affects the solvent exposure of Trp. The reaction in this conditions yielded essentially the same products - Fig. 5.17.

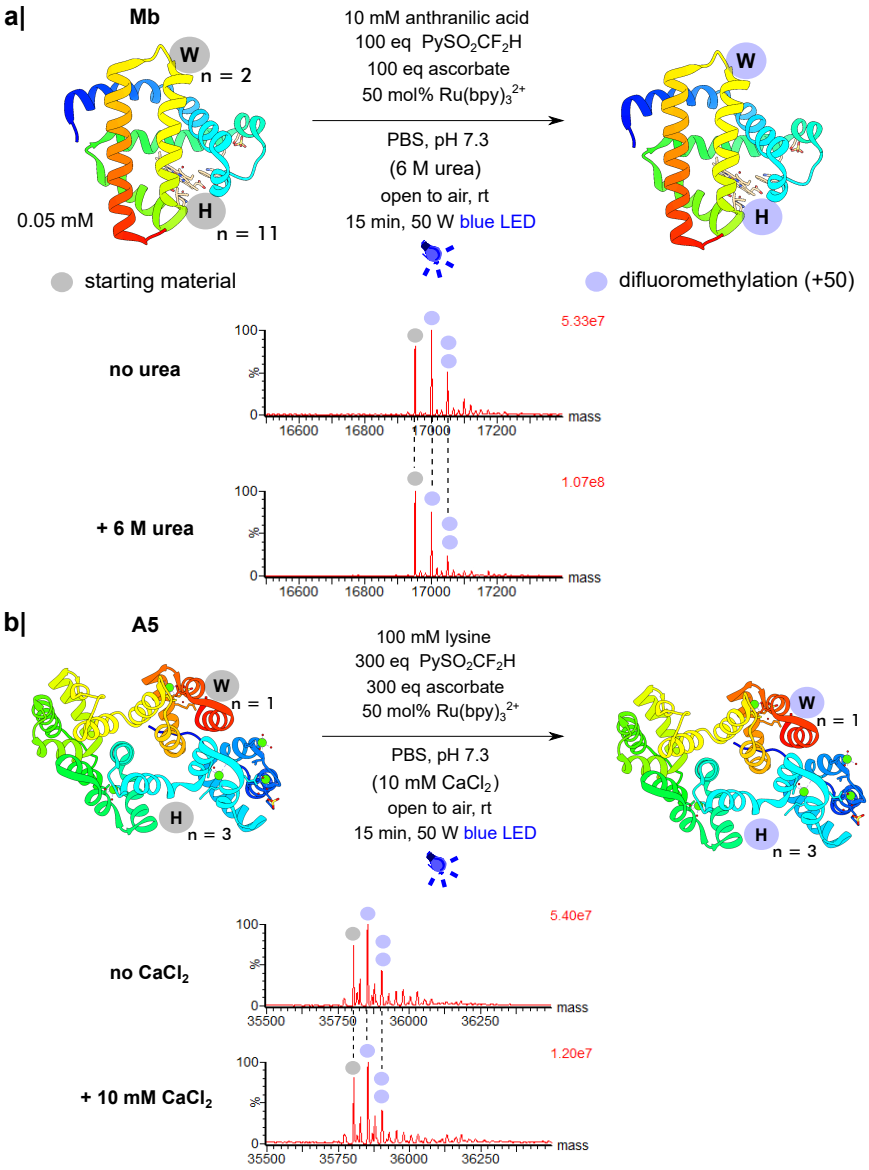


Fig. 5.17: Comparison of protein difluoromethylation under native and denaturing condition. **a**| LC-MS analysis of urea denatured myoglobin difluoromethylation indicated lower efficiency of the reaction, indicating possible interference of the chaotropic agent. **b**| LC-MS analysis of annexin V difluoromethylated in the presence of calcium revealed no changes to the modification pattern.

These results highlighted that $\bullet CF_2H$ had little steric requirements and a possible preference for the hydrophobic environment - an expected conclusion, based on the previous work on trifluoromethylation.

Finally, this hypothesis was further supported by the examination of LC-MS/MS data for the trypsinised difluoromethylated lysozyme sample - Appendix (Fig. 5.42). The C(O)H (originating from hydrolysis) modification was found to be statistically distributed across 5 Trp residues (out of 6). Importantly, the N-terminus and other aromatic residues were found to be only residually modified (as indicated by negligible relative intensities).

5.4.2 Hydrolysis

Having established a robust method for difluoromethylation on a number of proteins, the hydrolysis of the difluoromethyl tag was investigated - initially for myoglobin. Monitoring of the difluoromethylated protein in the reaction mixture quickly revealed that the transformation did not proceed according to the kinetics profile observed for the model peptide - Fig. 5.18. Instead, a very slow reaction was observed, accompanied with partial protein precipitation. The stability of the biomolecule could be improved by the removal of reactants (via buffer exchange), but this did not impact the formation of hydrolytic products, as samples analysed after incubation for 4 days showed minimal changes (data not shown). This suggested that solvent accessibility plays a role in stability of the CF_2H tag - both Trp residues in myoglobin are greatly shielded from the aqueous environment (see Tab. 2.2), explicitly required for the hydrolysis to occur.

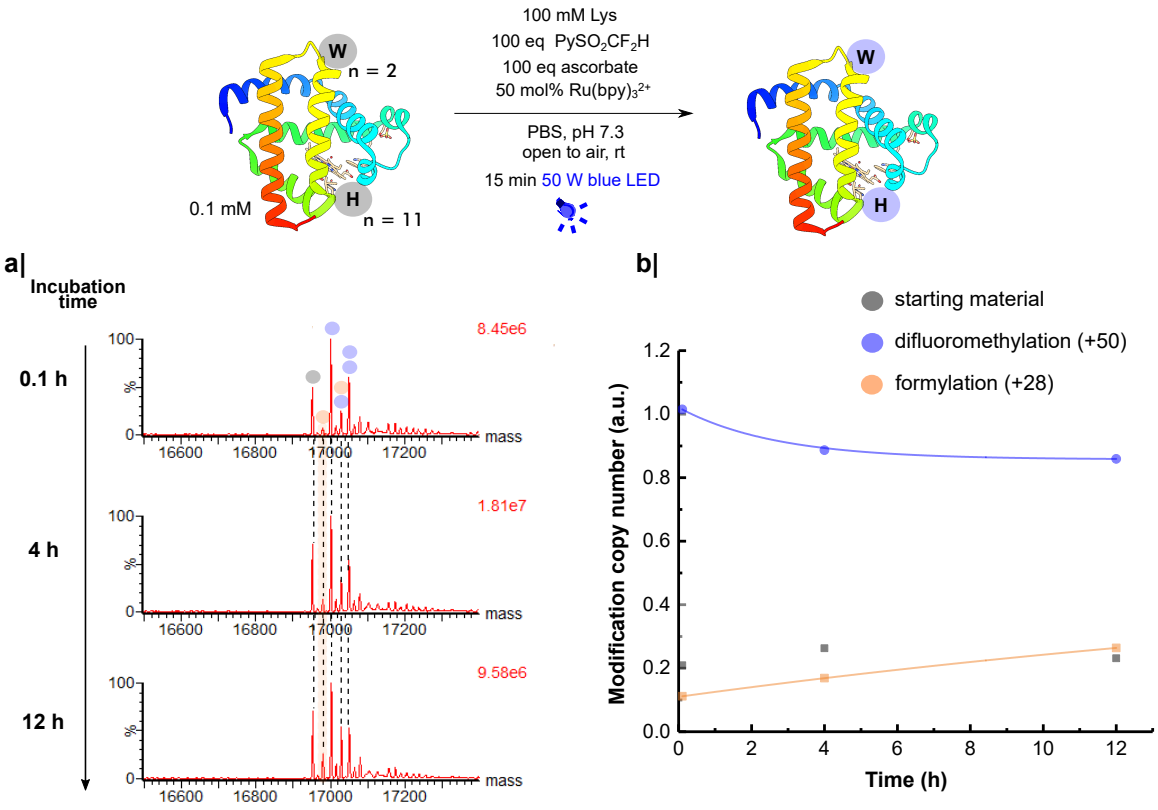


Fig. 5.18: LC-MS monitoring of hydrolysis for difluoromethylated myoglobin sample under native conditions. **a**| Deconvoluted mass spectra of the reaction mixtures at different timepoints. **b**| Quantification of different modification copy-numbers in time.

To test this hypothesis, a modified protein sample was incubated under denaturing conditions inducing the full unfolding of the protein. Use of high urea concentration (6 M) and elevated temperature (37 °C) allowed for the significant formation of hydrolytic products after incubation for 4 h - Fig. 5.19a; however, such conditions also promoted the carbamylation reaction, driven by the decomposition of urea[374]. This was evidenced by the formation of characteristic products by LC-MS. Although, the undesired carbamylation could be suppressed by use of a competitive amine - Fig. 5.19b, instead, a more stable chaotropic agent - guanidinium was investigated.

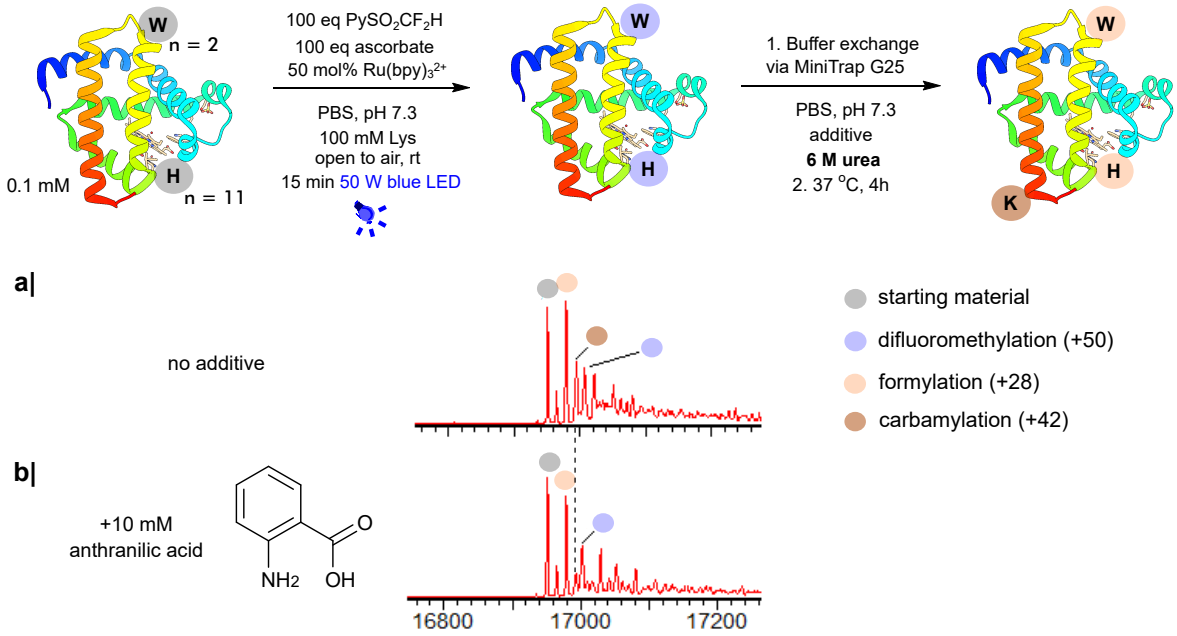


Fig. 5.19: Incubation of difluoromethylated myoglobin sample in a buffer with urea at elevated temperature. **a**| Mass spectrum of the reaction mixture after 4 h of incubation showing formation of new products (brown) in addition to expected hydrolysis (peach). **b**| The formation of carbamylation the product was not observed during the incubation in the presence of a sacrificial amine.

Using guanidine based denaturing strategy was also proved to be successful, with no additional side products being formed - Fig. 5.20. The formation of the formylated species was tracked over time, and complete conversion was observed after overnight incubation at 37 °C. These results further confirmed the need for protein denaturation to achieve the full hydrolysis of the difluoromethyl tag. The established conditions (14 h incubation in guanidinium buffer at 37 °C) will be referred to as standard conditions for hydrolysis.

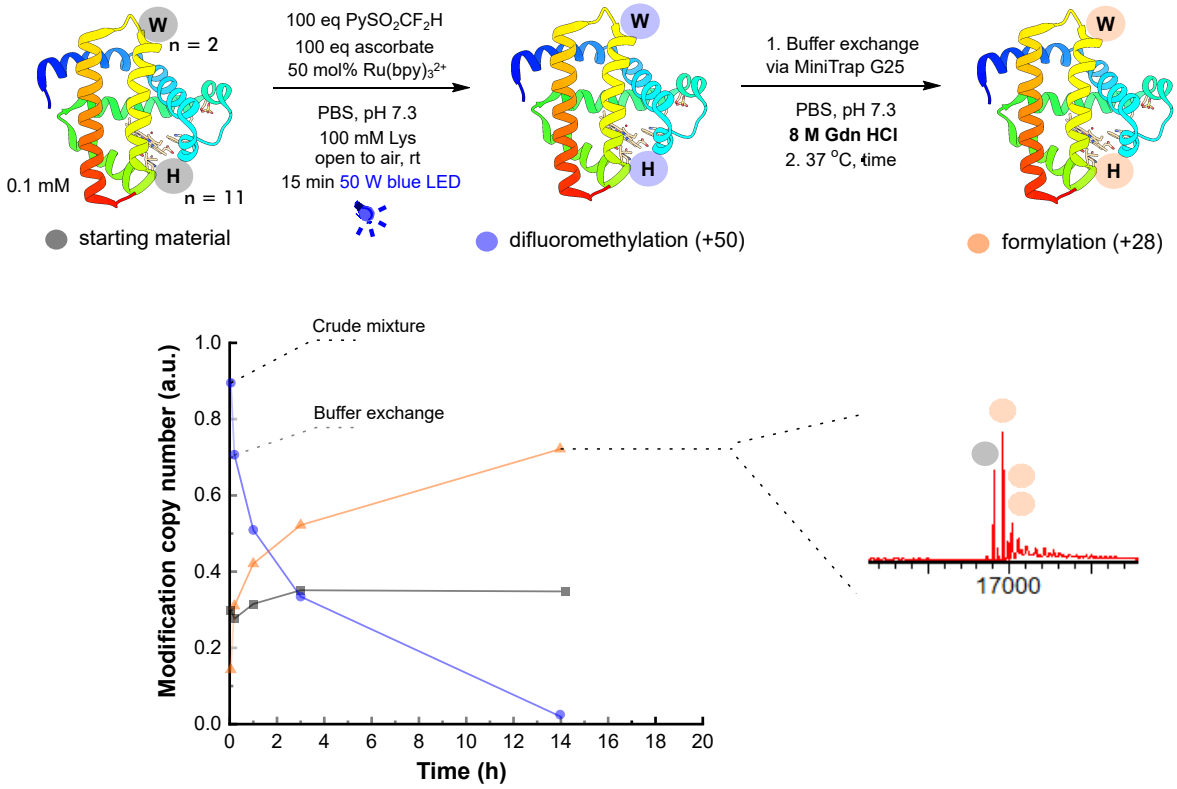


Fig. 5.20: Incubation of difluoromethylated myoglobin sample in a buffer with guanidinium at elevated temperature. The copy number of different modifications was tracked over time and revealed the gradual disappearance of the CF_2H modification, in favour of hydrolytic products. Mass spectrum (right) of the final timepoint showed the singly and doubly formylated products of expected mass.

To further prove that the observed formylation product constituted C-formylation, the borodeuteride reduction assay was carried out (see Fig. 5.10). Optimisation of the conditions (time, reagent excess) was required to reach a compromise between protein stability and suitable reactivity (not shown). Careful examination of the mass spectra was performed to determine accurate masses of the formylated and reduced protein sample. The analysis yielded the mass of the reduced peak (Fig. 5.21 - green shade) as $MW_{found} = 16982.2 \pm 0.1$, compared with $MW_{calc} = 16982.5$, and $MW_{found} = 16979.4 \pm 0.1$ for the formylated sample (Fig. 5.21 - peach shade) with $MW_{calc} =$

16979.5. This data allowed to determine that (92 ± 3)%, of the protein could be reduced, and hence had the modification was localised within Trp (and possibly His) as the C-aldehyde products.

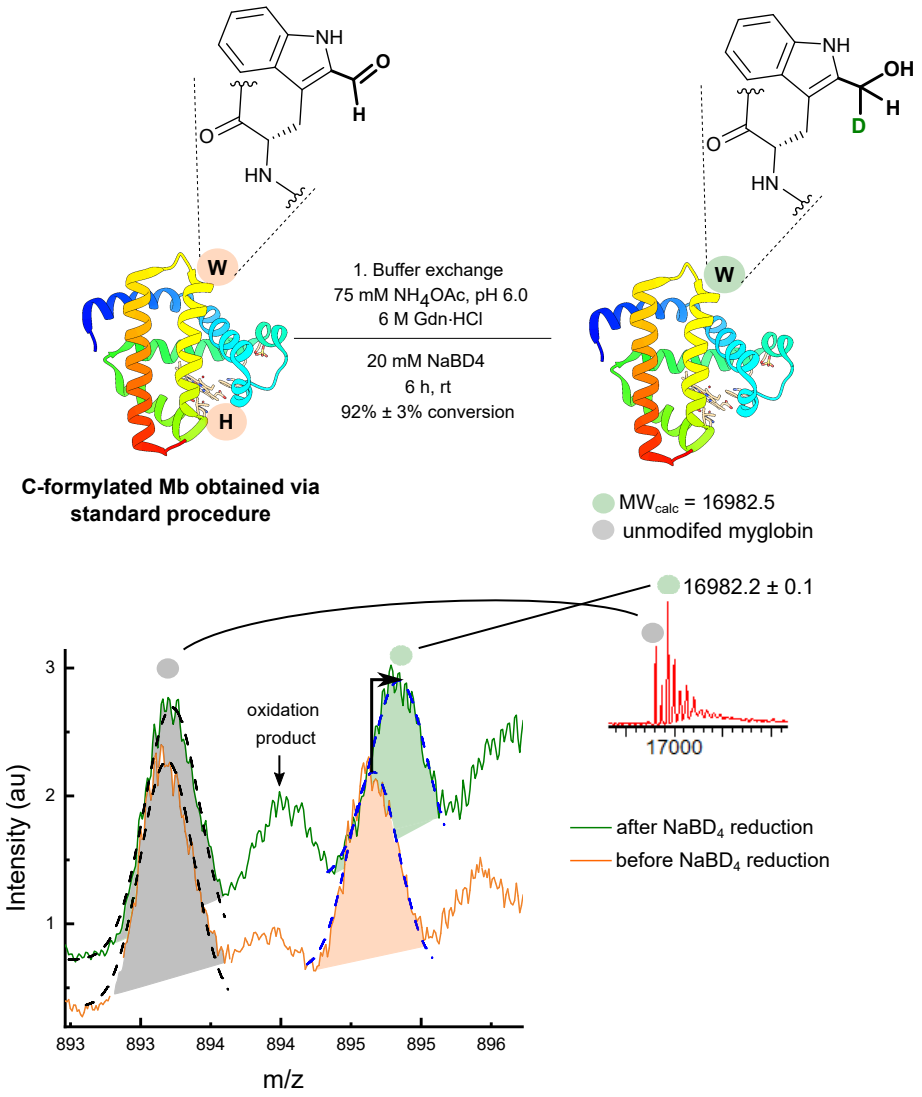


Fig. 5.21: The borodeuteride reduction assay for a myoglobin sample obtained via the standard difluoromethylation/hydrolysis sequence. To achieve a precise analysis, the raw ion series prior to deconvolution was analysed. Here, a close view on peaks for the +19 charged state of formylated myoglobin before (peach line) and after the reduction with $NaBD_4$ (green line). The peaks were fitted with a Gaussian function[375] (dashed lines, Origin v9.0) and their position was correlated with the protein molecular mass using the starting material peak as the reference (black shade).

This calculation was based on the fact that the discrepancy in the mass was attributed to incomplete conversion to the reduced product (due to the background residual N-formylation). Importantly, this proposed procedure constituted a new method for the introduction of a non-exchangeable deuterium atom into proteins.

Next, a sample of myoglobin, which was difluoromethylated under fortified conditions (300 eq of the radical precursor), was subjected the the hydrolysis conditions - Fig. 5.22. Predictably, this revealed that the stoichiometry of the reagents played a crucial role in maintaining Trp selectivity.

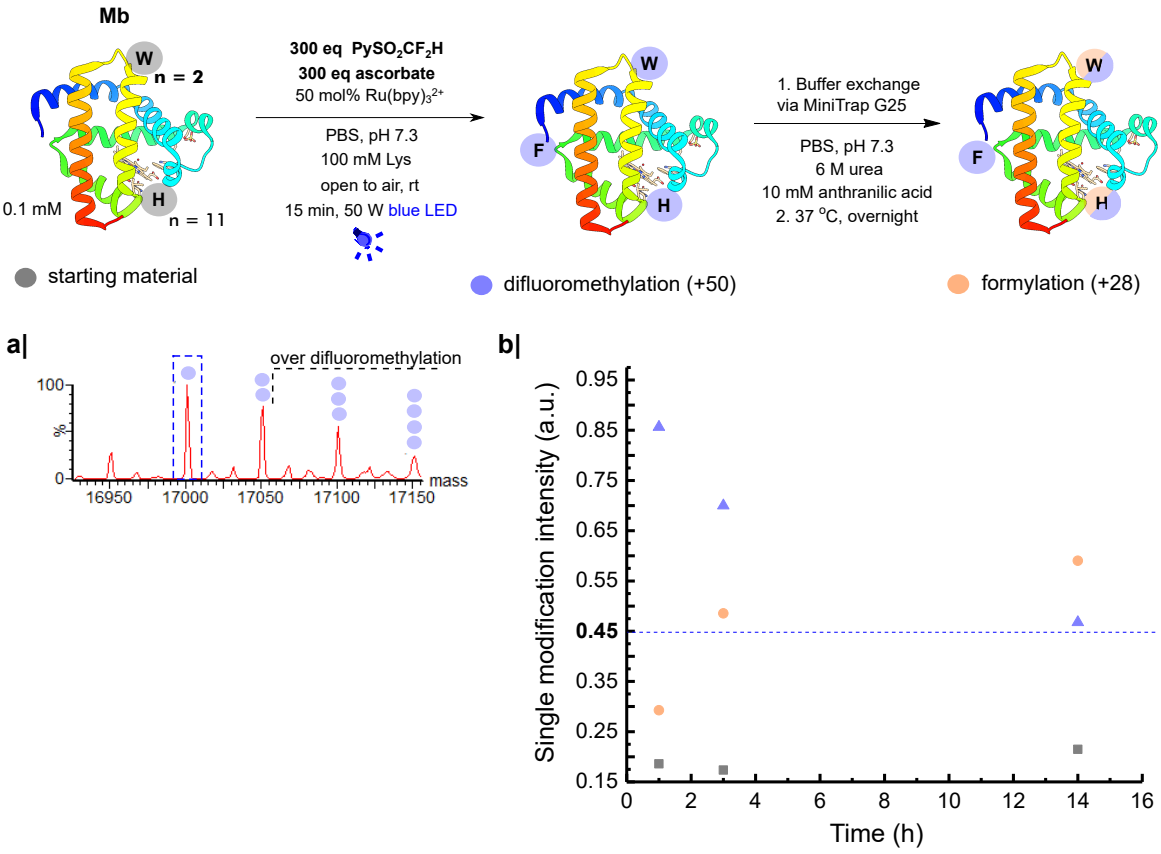


Fig. 5.22: Difluoromethylation of myoglobin under fortified conditions and denaturing hydrolysis. **a**| Mass spectrum of the crude reaction mixture **b**| LC-MS quantification of a single difluoromethylation and formylation modifications over time as well as the starting material (serving as an internal control of presumed constant intensity).

LC-MS analysis of the crude reaction mixture revealed multiple difluoromethylated products, over the expected Trp single addition products (> 2), clearly indicating multiple modification of a single Trp residue and/or Phe and His reactivity - Fig. 5.22a. In line with this hypothesis, when subjected to the standard hydrolysis conditions and analysed after the incubation, the CF_2H groups disappeared only partially - Fig. 5.22b. Therefore, such a denaturating hydrolytic assay could be used for establishing the optimal difluoromethylation conditions for each protein substrate, where the complete hydrolysis would indicate a full Trp selectivity, in agreement with the proposed modification of C2 positions of Trp residues.

5.4.3 Conjugation of aldehyde tagged proteins

After the successful translation of the difluoromethylation/hydrolysis sequence to protein substrates, attention was turned to evaluation of possible applications of the aldehyde tagged proteins with a unique Trp positioning. Initially, the potential of this aldehyde tag to perform the standard bioconjugation chemistry was assessed. One of the most common modifications of carbonyl groups is achieved with α-nucleophiles, due to their biocompatibility and unprecedented operational simplicity[376]. Considering that both hydrolysis and such a labelling step proceed favourably under similar conditions (prolonged denaturing incubation at elevated temperatures), and to simplify the procedure, a one-pot reaction was carried out. Only in one case was it observed that the presence of a nucleophilic catalyst was instrumental for the formation of the product (tosylhydrazone - Fig. 5.23), while other α-nucleophiles substrates did not require any additives - Fig. 5.24.

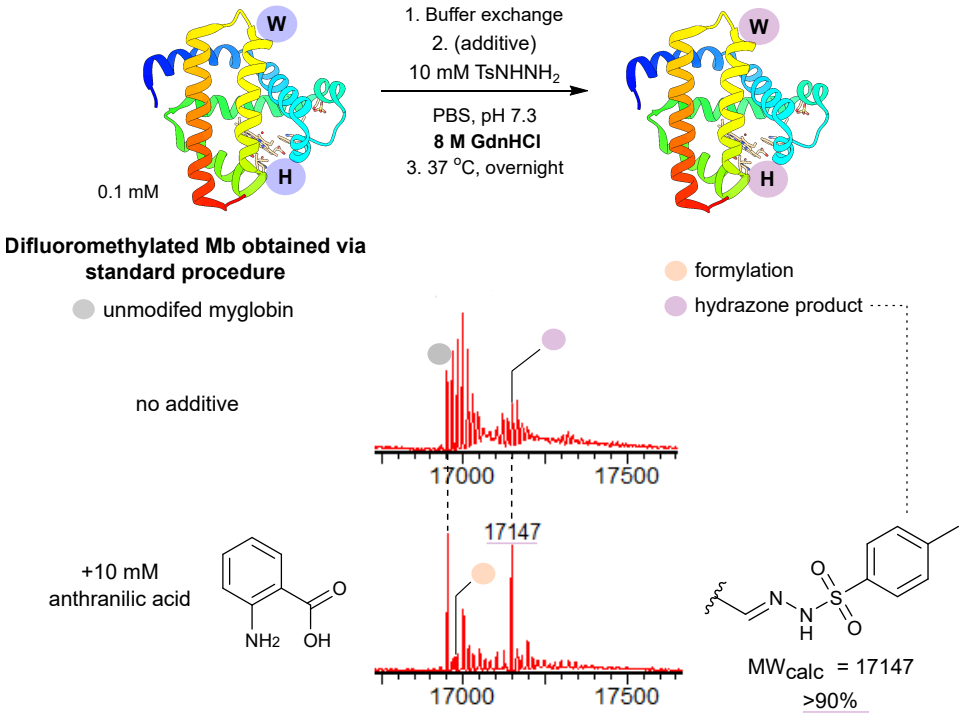


Fig. 5.23: One-pot hydrolysis/labelling sequence of difluoromethylated myoglobin using *p*-toluenesulfonyl hydrazide ($TsNHNH_2$), which required the nucleophilic catalyst for the product formation. Only in the presence of anthranilic acid a good mass spectrum with clear peaks was obtained.

Such coupled hydrolysis/labelling transformations resulted in efficient bioconjugation to a range of substrates (hydroxylamines, hydrazines) in over 85% conversion, as quantified relative to the remaining aldehyde - Fig. 5.24. Reactions with this Trp aldehyde were characterised by good kinetics, as overnight incubations with a 100x excess of the reagents were sufficient to obtain near complete conversions. Only the nitrone product, derived from *N-tert*-butyl hydroxylamine, was obtained in lower efficiency under studied conditions, presumably due to the steric hindrance of this nucleophile.

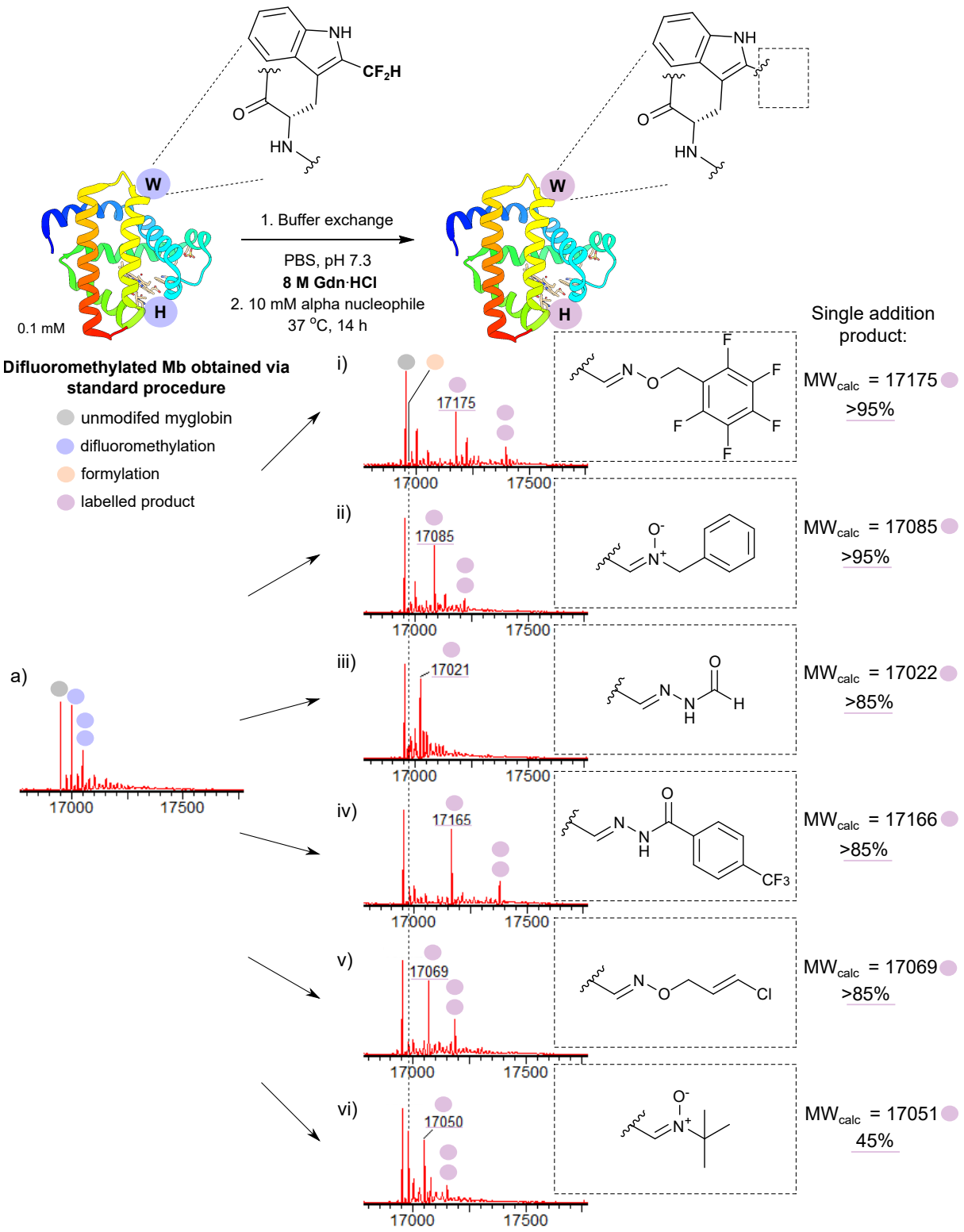


Fig. 5.24: One-pot hydrolysis/labelling of difluoromethylated myoglobin with alpha-nucleophiles. a) Mass spectrum of protein sample subjected to hydrolysis in the presence of labelling reagent. i)- vi) Mass spectra of reaction mixtures after an overnight incubation giving products of a shown structure as well as corresponding conversion.

Following this general procedure, it was possible to append payloads of function, such as: a preferred CF_3 tagged ^{19}F NMR probe - Fig. 5.25a, or an affinity tag - Fig. 5.25b. Successful conjugations were confirmed by the measurement of a relevant NMR spectrum and a positive response to an anti-biotin Western blot, respectively.

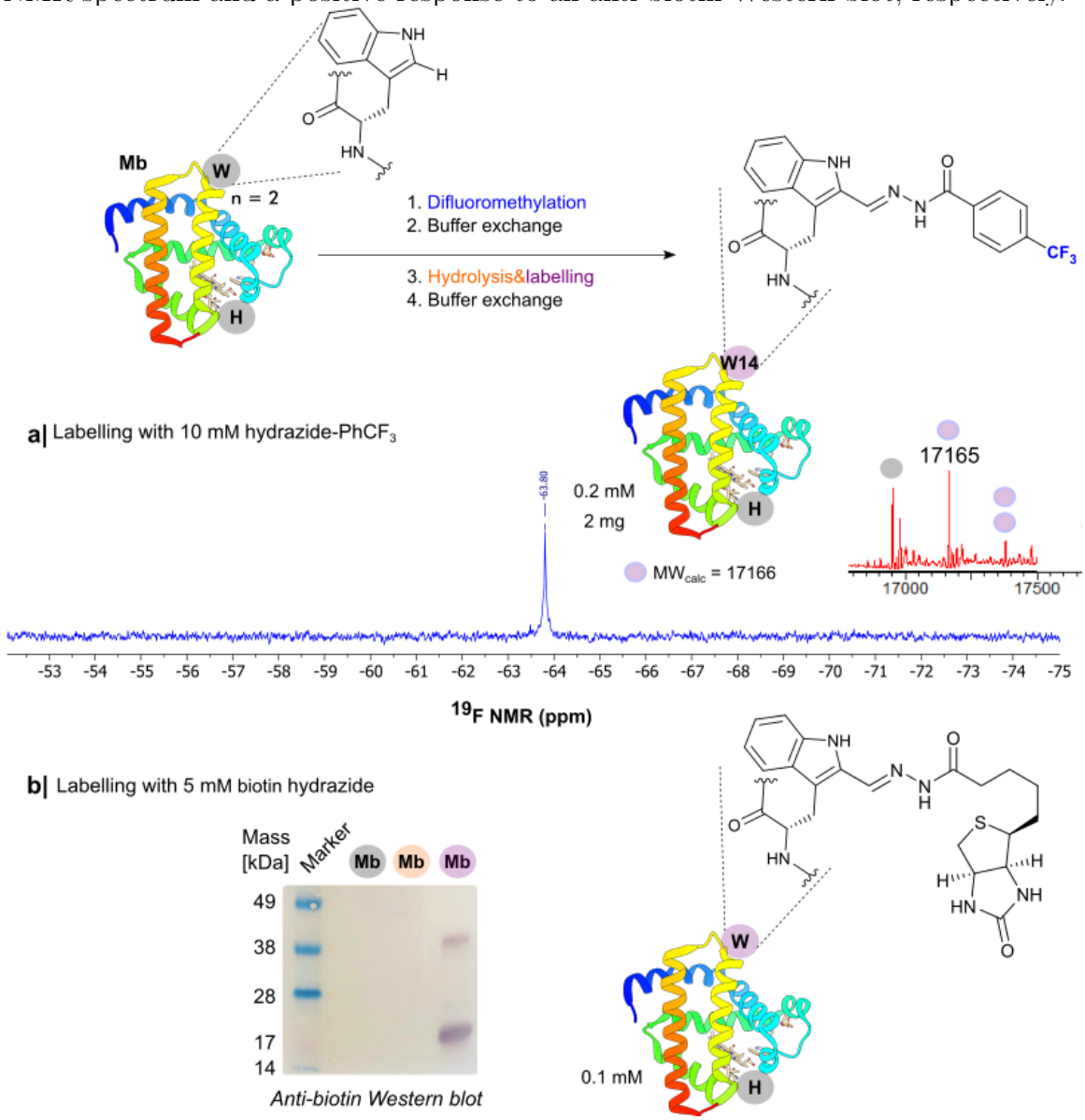


Fig. 5.25: One-pot hydrolysis/labelling with alpha-nucleophiles conveying function. **a**| Mass and ^{19}F NMR spectra of labelled myoglobin with the formed product structure. **b**| Anti-biotin Western blot with the band visible only for labelled myoglobin (no signal for the control formylated or unmodified protein). The less intense upper band was present in the commercial starting material (as revealed by SDS-PAGE, possibly the protein dimer).

Under the proposed conditions, labelling of myoglobin mostly yielded species with a single copy number of the conjugate product. In order to determine its specific location, a hydrazone tagged protein sample was digested with trypsin and peptides analysed in LC-MS/MS experiment. Peptide mapping revealed that the modification could be found on both Trp as well as several His residues - Appendix (Fig. 5.43). However, a careful semi-quantitative analysis of peptide relative conversions (taking into account signal intensity for the modified versus unmodified peptides) clearly pointed to Trp14 as the main site of the modification - Fig. 5.26.

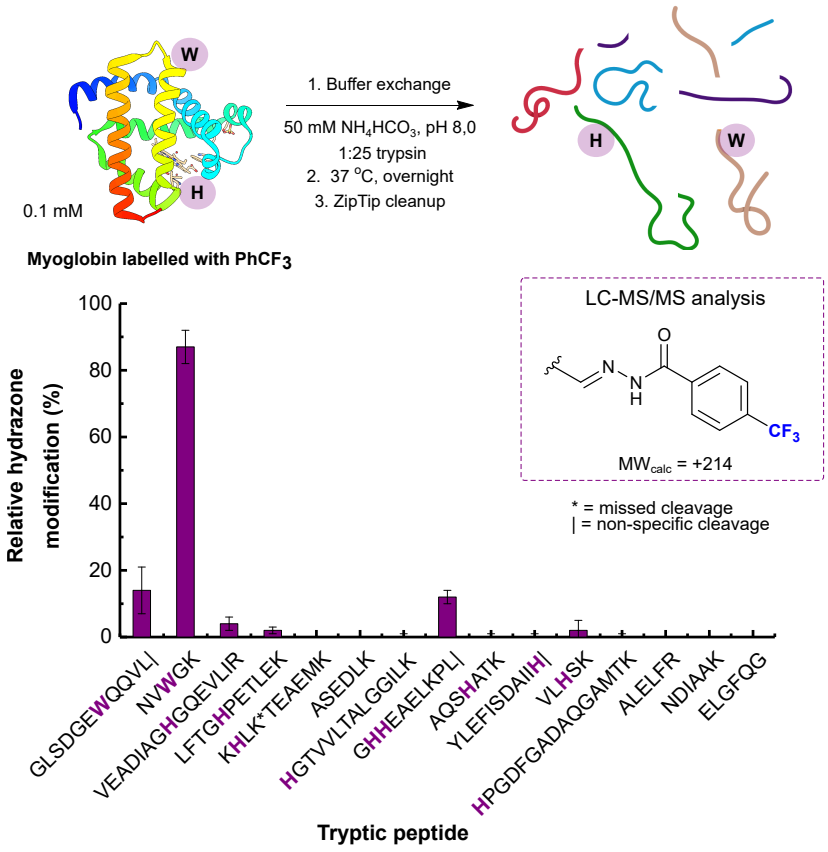


Fig. 5.26: LC-MS/MS label free quantification ($n = 4$) of the hydrazone modification within tryptic peptides from myoglobin modified under the standard sequence. The tryptic peptide coverage and cleavage patterns were reproducible in all experiments, reporting on 2/2 Trp and 10/11 His residues.

To verify that the proposed modification protocol could be easily extended to different protein substrates, biotinylation was performed for a range of different biomolecules including enzymes (thrombin, chymotrypsin and horseradish peroxidase) and proteins with specific functions (lactalbumin, annexin A5 and signal recognition nucleoprotein). In all of the a cases successful conjugation was confirmed via anti-biotin dot blot - Fig. 5.27.

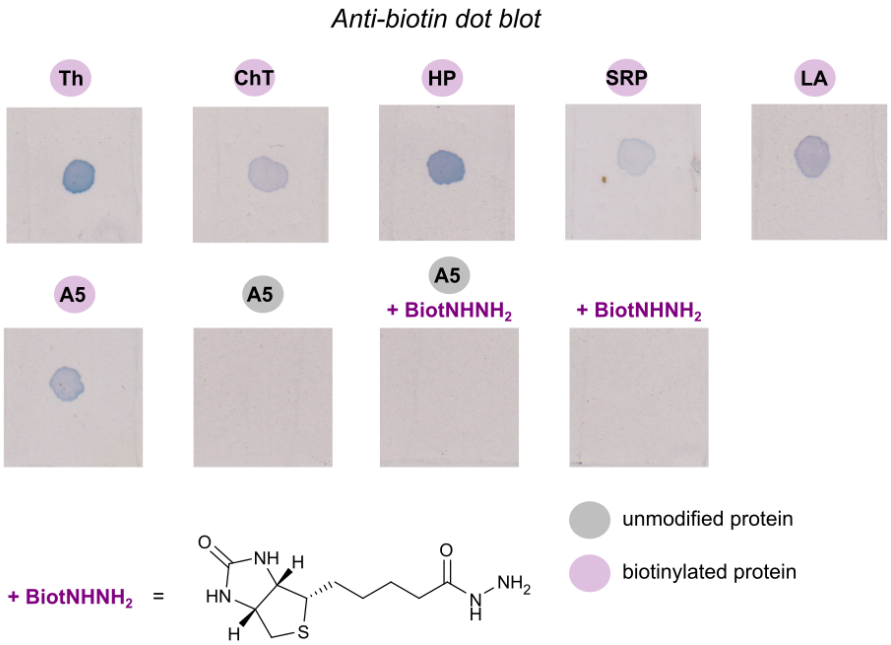


Fig. 5.27: Anti-biotin dot blot for a range of modified proteins labelled with biotin hydrazide under standard conditions (Th = bovine thrombin, ChT = bovine chymotrypsin, HP = horseradish peroxidase, SRP = His-tagged human signal recognition particle 19 kDa, LA = bovine lactalbumin, A5 = wild type chimpanzee annexin V) along with controls: +BiotNHNH$_2$ in which the labelling reagent: biotin hydrazide was spotted. In all cases, 1 μL of a protein solution was applied to the membrane. The proteins were prepared in a range of concentration (20 – 100) μM in PBS buffer, while the hydrazide was used as 5 mM DMSO solution.

It is well known that hydrazone/oxime products are reversible in aqueous solution[376], so in order to create a stable linkage, an alternative strategy using reductive

amination, which generates a stable amine linked conjugate, was explored. In an analogous manner, a one-step hydrolysis/labelling sequence was investigated with a range of aromatic amines (displaying better nucleophilicity in the preferred neutral aqueous buffers[377]) and $NaBH_3CN$ for the reduction of intermediate imines. This procedure resulted in the formation of the expected products from different anilines in $> 80\%$ conversion - Fig. 5.28. Interestingly, the efficiency of the reaction was found to be somewhat reduced with alkyl amines, requiring a much higher excess of the reagent, which was possibly correlated with their pK_a and the reduced population of deprotonated amine in the neutral buffer. However, it was found that for a range of tested substrates, a standard procedure utilising 20 mM amine and 50 mM $NaCNBH_3$ provided the product in good conversion; however for aliphatic amines, and other systems, optimisation of pH and stoichiometry may be required. In summary, the hydrolysed CF_2H tag was demonstrated to be a versatile, reactive handle, allowing ready access to variety of Trp mimics upon reactions with nucleophiles.

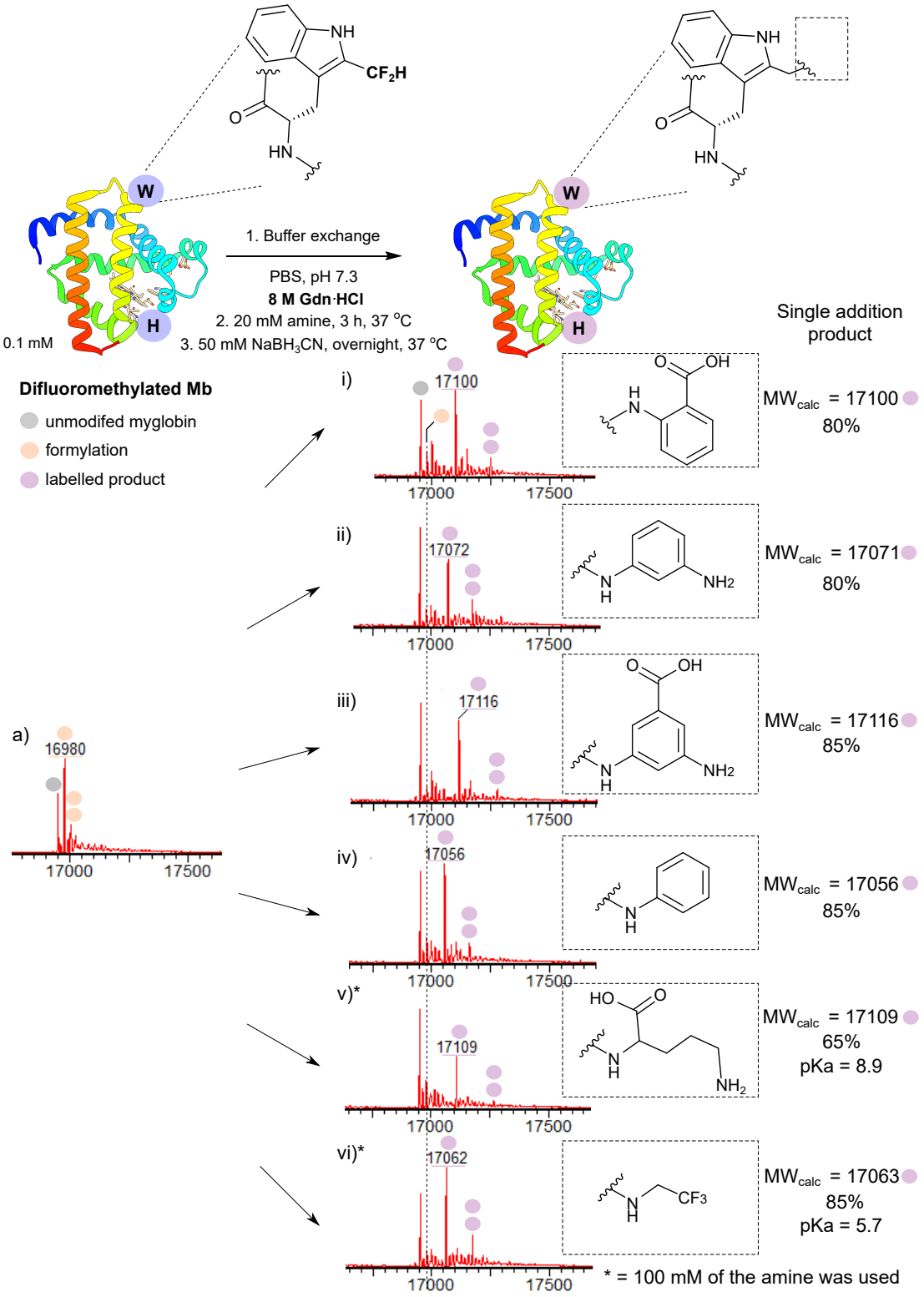


Fig. 5.28: One-pot hydrolysis/labelling via reductive amination for the creation of stable products. a) Mass spectrum of the non $NaBH_3CN$ treated sample, which was subjected to hydrolysis in the presence of labelling reagents. i)- vi) Mass spectra of reaction mixtures after an overnight incubation with reagents, the structure of the product and conversion.

5.4.4 Fluorescent properties of aldehyde tagged proteins

Tryptophan is a primary natural fluorophore that has been used extensively in biophysical studies of proteins[378]. The use of its intrinsic fluorescence is particularly valuable, as no perturbations accompanying the introduction of fluorescent tags are required, preserving the native interactions and function of proteins. The chemistry described herein targets the core aromatic structure of this residue; extension of the conjugated system at the C2 position of indole-like structures is a known synthetic strategy for the construction of potent fluorescent dyes and indicators[379].

Modification of Trp, in order to shift the fluorescence towards the visible region, has been proposed in order to facilitate protein spectroscopy. Importantly, a C2-aldehyde on Trp was was shown to be a suitable, minimally sized structural motif which indeed shifts the fluorescence in a small molecule model[380]. Initial synthetic efforts to realise this idea on proteins were based on a photochemical reaction between chloroform (or other trihalocompounds) and the indole ring excited by UV light[381,382]. While this approach achieves the desired outcome of products with a red-shifted emission spectrum, a large heterogeneity of the resulting species and the damaging conditions of the reaction meant that this procedure found applications in the analysis of proteins separated with SDS-PAGE - in so called 'stain-free' visualisation[383]. Considering that the protocol developed here enables access to highly selective C-formylation, it was proposed that the modified proteins would display the improved intrinsic fluorescent properties.

Proteins separated with SDS-PAGE were transilluminated with broad-range ultra-

violet light and their emission was detected at wavelengths above that of native Trp (>480 nm)[384], using an appropriate cut-off filter - Fig. 5.29a. Pleasingly, the bands were only visible for samples subjected to the difluoromethylation/hydrolysis sequence.

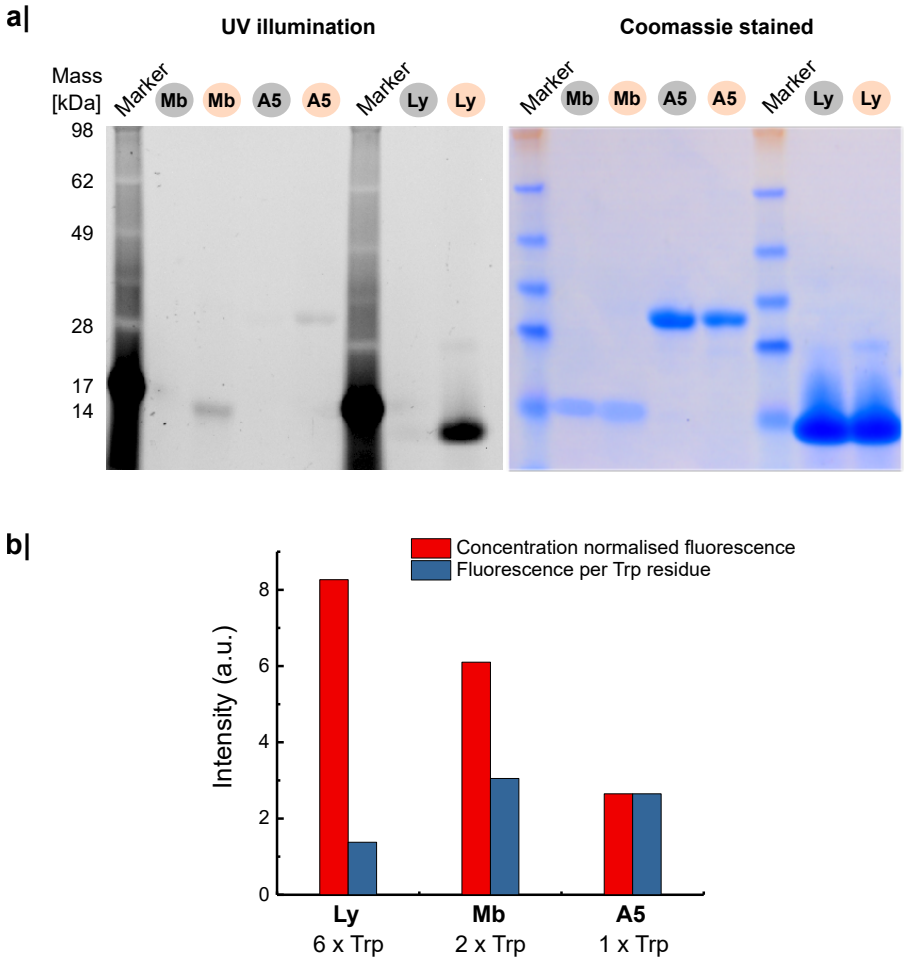


Fig. 5.29: Analysis of fluorescence properties of the proteins after the difluoromethylation/hydrolysis sequence. **a**| Comparison of SDS-PAGE results with the gel visualised by UV light transillumination or staining with InstantBlue Coomassie Protein stain (Mb = myoglobin - 3 μg per lane, A5 = annexin 5 - 6 μg per lane, Ly = lysozyme - 20 μg per lane, both lanes were loaded with the same amount of protein). **b**| Quantification of concentration weighted to the relative intensity of the bands (densitometry analysis with ImageJ (1.52a) software - red bars), related with the number of Trp residues (blue bars).

The quantification of the concentration normalised fluorescence, revealed the expected

correlation with the number of modified Trp residues - Fig. 5.29b, confirming lysozyme (6 reactive Trp sites) as the most intense fluorescent species. This protein also displayed the *smallest* contributions from individual residues, which was in agreement with results from intact LC-MS analyses, showing much lower conversions to high copy number difluoromethylated species (average 1.75 copy number, 29% of the maximal Trp content based conversion), as compared with modified myoglobin (0.84 copy number, 42% Trp conversion) or annexin V (0.40 copy number, 40% Trp conversion).

To further characterise the modified proteins, the actual emission spectra were recorded and compared with that of native proteins - Fig. 5.30. As expected, the highest relative increase of long wavelength fluorescence (with local maximum of approximately (456 $\pm$ 5) nm) was observed for lysozyme, in which the high relative content of the modified residues was hypothesised to be responsible for this feature. For single Trp proteins: annexin or signal recognition particle proteins, this was less noticeable, possibly suggesting a lower quantum yield for fluorescence[380] of modified residues or some additional quenching mechanisms or interactions.

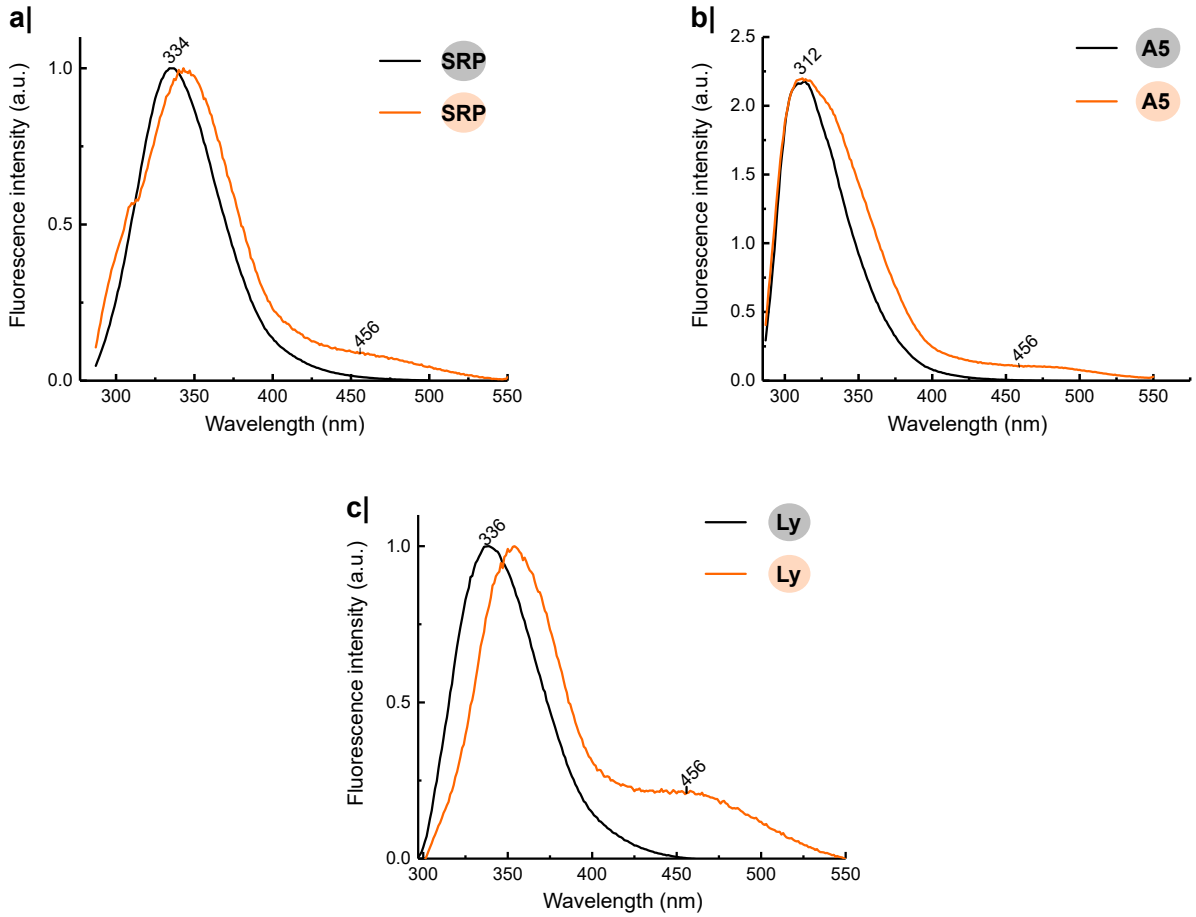


Fig. 5.30: A comparison of the normalised emission spectra (λ_{ex} = 280 nm) for different proteins after the difluoromethylation/hydrolysis sequence. The proteins were refolded into PBS pH 7.3 before the measurement - **a**| 10 μM signal recognition particle (SRP), **b**| 15 μM annexin V, **c**| 20 μM lysozyme. Note the fluorescence profile for annexin V, dominated by the Tyr residues emission (12 Tyr vs 1 Trp).

5.5 Discussion and conclusions

In conclusion, the study of photoredox catalysed difluoromethylation of native residues led to the development of a new protocol for the introduction of uniquely positioned aromatic aldehydes into proteins. Firstly, a reactivity hypothesis regarding hydrolytic instability of the difluoromethyl modification within heteroaromatic amino acids was

confirmed on a small molecule model, which also helped to confirm the dominant Trp selectivity under the proposed conditions. The radical generation system was based on a fluoroalkylated pyridyl sulfone as opposed to a sulfinate, used previously for trifluoromethylation. Interestingly, difluoromethyl sulfinate was also demonstrated to be a light activated radical precursors suitable for the heteroaromatic substrates modification (including 3-substituted indoles), however not within proteins[385]. The reason why this strategy was not pursued for the protein modification here, is that such a protocol is predicted to suffer from the same problem of the sulfinate oxidative decomposition in aqueous buffers (leading to formation of N-reactive species), what was observed in the case of difluoroethylation (Chapter 4). In further support of this, various *alkyl* sulfinates, which do not follow this reactive pathway, have been used for photoredox modification of dehydroalanine in protein substrates[156]. Here, using the pyridyl sulfone - $PySO_2CF_2H$, the difluoroalkyl radical could be accessed with limited side reactivity that was easy to counter by with competitive amines.

The recently disclosed difluoromethylation of indoles in a similar photoredox driven radical process has also supported the C2 selectivity[386]. Interestingly, the work where this type of catalysis was used for difluoromethyl sulfinate activation for modification of Trp derivatives, has highlighted the role of oxygen as a possible terminal oxidant[385]. Although oxygen often is detrimental to radical chemistries, especially in the presence of fragile proteins, in this thesis little oxidative damage was observed, presumably due to its role in the mechanism (terminal oxidant) as well as the scavenging role of ascorbate. In fact, when protein difluoromethylation was performed in an O_2-free glovebox the reaction resulted in lower yields (data not shown).

Studies on peptides provided further insights into the mechanism of the reaction, allowed optimisation of the conditions, confirmed the observed trends and identified possible side reactions. These results facilitated the development of an efficient protein difluoromethylation protocol, in which the use of a sacrificial amine in the buffer was found to efficiently protect the proteins against the undesired N-formylation. Other side reactions came from the fact of performing a redox intensive transformation in aerated (oxygenated) solution (such as exposed Met and Cys oxidation). However, the operational ease and the access to broad range of protein substrates offset that drawbacks.

The protein difluoromethylation was characterised in-depth, elucidating its unique features such as its ability to access the buried Trp residues and stoichiometry driven selectivity. The investigation of hydrolysis conclusively confirmed preservation of the modified protein fold, through observation of the sterically inhibited hydrolysis of the CF_2H tag. This revealed an exciting possibility of the difluoromethyl group as a shielded aldehyde tag, which could be revealed only under specific conditions (engagement in an interaction, degradation etc.). Such an approach was envisioned before, as a countermeasure to the potential toxicity of this functional group[387]. Trp residues often participate in protein–protein (ligand) interactions or appear at/near protein membrane interfaces[271], which notably renders Trp with an electrophilic tag an exciting handle for the cross-linking experiments and trapping of typically elusive protein-protein interactions.

Different denaturing protocols were evaluated to expose the C-formylated Trp residues and to demonstrate synthetic utility of the introduced aldehyde - the group

behaved excellently in common bioconjugation procedures with α-nucleophiles as well as also readily formed easily reduced imines with a variety of partners, allowing for the formation of a stable linkage and showing its strength as a reactive motif for he 'Tag and Modify' strategy - Fig. 5.31.

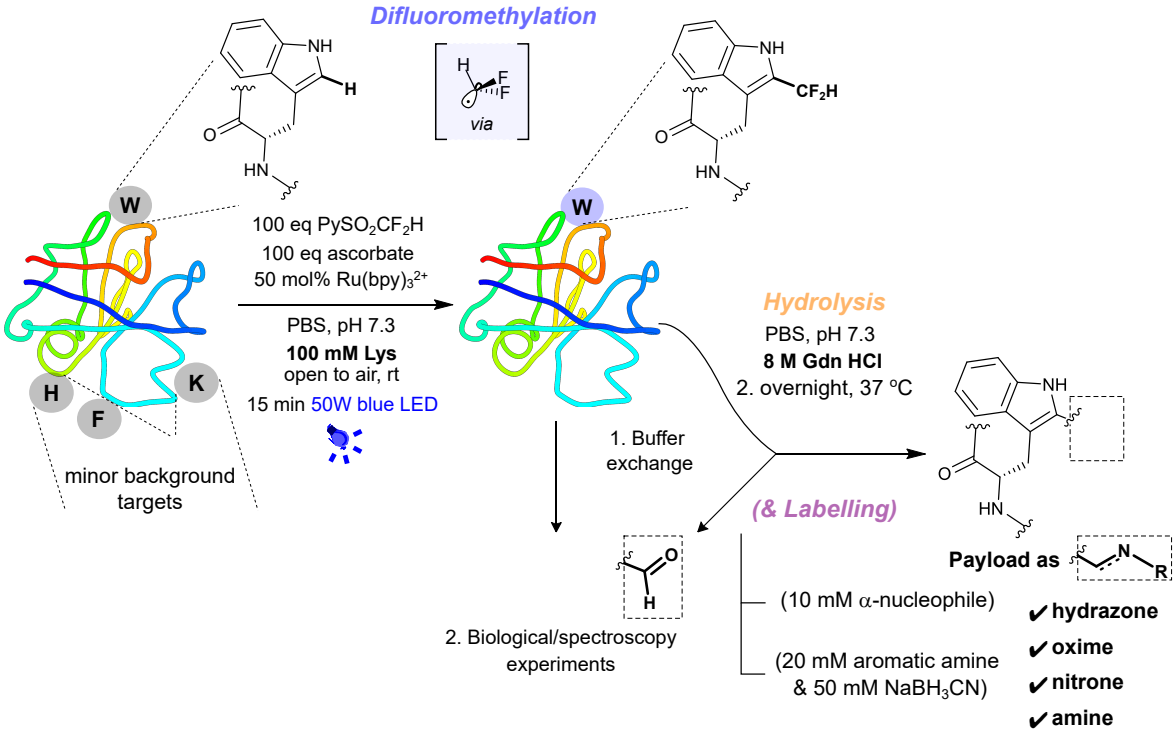


Fig. 5.31: Overview of the proposed standard conditions for photoredox difluoromethylation and coupled hydrolysis/labelling sequence for the Trp selective modification of proteins developed in this thesis. Proposed applications are based on the unique site of conjugation (Trp residues) and minimal size fluorescent tag.

This could be used not only for general probe attachment, but also could potentially to study the function of native, post-translationally C-modified-Trp (one of the most common being C2-mannosylation)[388] using close mimics. The developed protocol valuably provided previously non-existent tools to study this modification, as even though it was found almost three decades ago, only scarce findings have pro-

vided insights into its role in biology[389]. Required homogeneously modified products can be obtained by subjecting the aldehyde tagged proteins in a catch and release purification[390]. It is worth noting that the carbonyl tag, despite its already achieved tremendous versatility in protein chemistry, still attracts a special interest in the community as more and more novel procedures are being developed for its applications in the bioorthogonal labelling of peptides and proteins[391].

Furthermore, the extension of the fluorophore system through C-formylation was found to alter the fluorescent properties of the proteins, possibly enabling applications in imaging and different experiments in photobiology. With such a broad applications, the developed difluoromethylation/hydrolysis protocol is expected to be an enabling tool for further chemical biology experiments.

5.6 Experimental

For general information regarding the used experimental methods please refer to the relevant sections in Chapter 2 and 4. Additional proteins, not described in previous chapters, used in this study were sourced as followed:

Tab. 5.2: Identification of proteins used in the study and their source.

Abb.	**Name**	**Species**	**UNIPROT**	**Source**
Th	plasma thrombin	*Bos taurus*	THRB_BOVIN	Sigma-Aldrich (T4648)
ChT	pancreas α-chymotrypsin	*Bos taurus*	CTRA_BOVIN	Sigma-Aldrich (C4129)
HP	horseradish peroxidase	*Armoracia rusticana*	PER1A_ARMRU	Sigma-Aldrich (77332)
LA	milk lactalbumin	*Bos taurus*	LALBA_BOVIN	Sigma-Aldrich (L5385)

5.6.1 Experimental procedures

LC-MS assay for quantification of amino acid reactivity

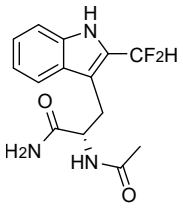
Stock solutions of amino acids (Trp, His, Phe or Tyr) were prepared in PBS buffer (7.5 mM). For the competitive reaction (with all four amino acids) different amino acids stocks were combined (81 μL each), and to such a mixture $PySO_2CF_2H$ (difluoromethyl 2-pyridyl sulfone, 6.2 μL, 1 M in DMSO) was added, as well as sodium ascorbate (3.7 μL, 1.65 M in PBS buffer) and $[Ru(bpy)_3]^{2+}$ (9.3 μL, 3.3 mM in PBS). ^{15}N-labelled amino acid standard mix was prepared by 5x dilution with ddH_2O of *Cell Free Amino Acid Mixture-^{15}N (Sigma-Aldrich #767972).* This mixture was centrifuged before use to pellet insoluble residues (presumably cystine) present in the commercially delivered sample. An aliquot (25 μL) of the 4 AAs reaction mixture was mixed with ^{15}N-AA mix (25 μL) and analysed via LC-MS (sample before irradiation). The remaining reaction mixture (100 μL) was irradiated with blue LEDs (450 nm, 50 W) for 15 mins, after which an aliquot (25 μL) was mixed with ^{15}N-AA mix (25 μL) and analysed via LC-MS (sample after irradiation). Small molecule LC-MS quantification assay used the following elution profile: 100% A $\rightarrow$ (0 – 2.5) min, gradient to 50% B $\rightarrow$ (2.5 – 7) min, 95% B $\rightarrow$ (7.1– 8) min, 100% A $\rightarrow$ (8.1 – 10) min with C18 BEH Agilent column.

LC-UV assay for confirmation of Trp reactivity in the reaction mixture

The reaction mixture was prepared as described above; however, higher amounts of $PySO_2CF_2H$ and ascorbate were used (40 eq of each). Analysis was performed under different LC conditions (using C18 Gemini N column (Phenomenex) and linear 20

min gradient of MeCN in aqueous 0.1% formic acid) without addition of ^{15}N-AA mix and dilution of the samples.

NATA difluoromethylation



NATA (*N*-acetyltryptophanamide, 29.4 mg, 0.12 mmol), $PySO_2CF_2H$ (46.4 mg, 0.24 mmol), sodium ascorbate (47.5 mg, 0.24 mmol) and $[Ru(bpy)_3]^{2+}$ (0.9 mg, 1 mol%) were weighed into a glass vial and water (1 mL) was added. After mixing, DMSO (1 mL) was added to the resulting suspension which allowed full solubilisation of the reactants. The reaction mixture was irradiated with blue LEDs (450 nm, 50 W) for 6 h and lyophilised afterwards. The residual DMSO solution was applied to a packed silica column (12 g) and chromatographed using a CombiFlash system and $CHCl_3$ with MeOH gradient up to 10%. All fractions with absorbance at 280 nm were collected and solvent was removed in vacuo. Among those, C2-difluoromethylated product **TrpCF_2H** (11.0 mg, 30%) was isolated as a white powder. Other fractions contained other NATA regioisomers (isolated as mixture <10%) and oxidised NATA (<10%) as well as NATA starting material (∼30%). The identities of these products were assigned based on LC-MS analysis as no pure compounds were isolated.

^{1}H NMR (600 MHz, DMSO-d6) δ 11.5 (s, 1H, H1), 7.98 (d, J = 8.8 Hz, 1H, NH), 7.75 (d, J = 8.0 Hz, 1H, H4), 7.44 (s, 1H, NH_2), 7.36 (d,J = 8.2 Hz, 1H, H7), 7.26 (t, J = 53.5 Hz, 1H, H9), 7.19 (dd, J = 8.2, 6.9 Hz, 1H, H6), 7.08 (s, 1H, NH_2), 7.05 (ddd, J = 8.0, 6.9, 1.0 Hz, 1H, H5), 4.44 (ddd, J = 8.8, 8.4, 5.3 Hz, 1H, Hα), 3.22

(dd, J = 14.4, 5.3 Hz, 1H, Hβ), 2.95 (dd, J = 14.4, 8.4 Hz, 1H, Hβ), 1.75 (s, 3H, $COCH_3$),

^{13}**C NMR** (151 MHz, DMSO-d6) δ 174.1 ($CONH_2$), 170.7 ($HNOCCH_3$), 136.4 (C7a), 127.7 (t, J = 22.2 Hz, C2), 127.2 (C3a), 124.2 (C6), 120.3 (C4), 120.0 (C5), 112.7 (t, J = 7.1 Hz, C3), 111.6 (C7), 110.7 (t, J = 231.2 Hz, C8), 53.9 (Cα), 26.7 (Cβ), 22.9 ($NOCCH_3$),

^{19}**F NMR** (565 MHz, DMSO-d6) δ −107.89 (dd, J = 308.0, 53.5 Hz, 1F), −108.53 (dd, J = 53.5 Hz, 1F), due to the overlapping peaks, the elucidation of the coupling constants was achieved by measurement of the compound dissolved in $CDCl_3$,

(376 MHz, $CDCl_3$) δ −108.51 (dd, J = 308.0, 53.5 Hz, 1F), −110.82 (dd, J = 308.0, 53.5 Hz, 1F),

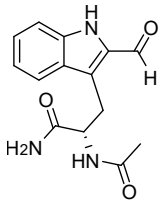
^{15}**N NMR** (51 MHz, DMSO-d6) δ 126.7 (N1), 123.0 ($HNOCCH_3$), 106.9 ($CONH_2$),

HRMS-ESI(+): $[M+Na]^+$: calcd: 318.1025, found: 318.0991,

$[M+H]^+$: calcd 296.1205, found: 296.1114.

TrpCF$_2$H hydrolysis to TrpC(O)H

TrpCF$_2$H (7.4 mg, 0.025 mmol) was dissolved in PBS buffer prepared in D_2O (0.5 mL). After an overnight incubation at rt without stirring, the sample was analysed by NMR and no starting material was observed, indicating a full conversion to the **TrpC(O)H** aldehyde product. The other detected product was fluoride anion (singlet c. (−122 ppm)).



^{1}**H NMR** (500 MHz, PBS D_2O) δ 9.87 (s, 1H, H9), 7.86 (dt, J = 8.2, 1.0 Hz, 1H,

H4), 7.54 (dt, J = 8.4, 0.9 Hz, 1H, H7), 7.50 (ddd, J = 8.3, 6.9, 1.1 Hz, 1H, H6), 7.27 (ddd, J = 8.1, 6.8, 1.0 Hz, 1H, H5), 4.71 (dd, J = 7.0, 7.5 Hz, 1H, Hα), 3.66 (dd, J = 14.4, 7.0 Hz, 1H, Hβ), 3.53 (dd, J = 14.4, 7.5 Hz, 1H, Hβ), 1.89 (s, 3H, CH_3),

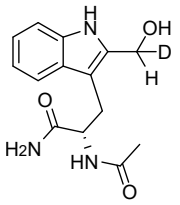
^{13}C NMR (126 MHz, PBS D_2O) δ 183.6 (C8), 175.2 ($CONH_2$), 173.7 ($NOCCH_3$), 137.9 (C7a), 127.9 (C6), 127.1 (C3a), 123.6 (C2), 121.2 (C4), 120.9 (C5), 112.9 (C7), 69.6, 54.2 (Cβ) 25.8 (Cα), 21.6 (CH_3),

HRMS-ESI(+): $[M+Na]^+$: found: 296.1103, calcd: 296.1006,

$[M+H]^+$: found: 274.1259 calcd 274.1186.

NATA-C(O)H reduction

To **TrpC(O)H** (6.8 mg, 0.025 mmol) dissolved in PBS buffer prepared in D_2O (0.5 mL), $NaBD_4$ (10.5 mg, 0.25 mmol) was added. The sample was incubated for 24 h and lyophilised afterwards. The crude residue was dissolved in DMSO-d6, filtered and analysed by NMR, which showed 90% conversion to the reduced alcohol **TrpCD(OH)H** that was not isolated (with the starting material as the only other species detectable by NMR).



^{1}H NMR (500 MHz, DMSO-d6) δ 10.89 (s, 1H, H1), 7.94 (d, J = 8.1 Hz, 1H, NH), 7.58 (d, J = 7.9 Hz, 1H, H4), 7.40 (s, 1H, NH_2), 7.27 (d, J = 8.0 Hz, 1H, H7), 7.01 (dd, J = 8.0, 6.9 Hz, 1H, H6), 6.97 (s, 1H, NH_2), 6.93 (dd, J = 7.9, 7.0 Hz, 1H, H5), 5.32 (broad d, J = 5.0 Hz, 1H, OH), 4.57 (broad d, J = 5.0 Hz, 1H, H9), 4.35 (td, J = 8.1, 5.7 Hz, 1H, Hα), 3.08 (dd, J = 14.3, 5.7 Hz, 1H, Hβ), 2.92 (dd, J = 14.3, 8.1 Hz, 1H, Hβ), 1.74 (s, 3H, $COCH_3$),

^{13}C NMR (126 MHz, DMSO-d6) δ 173.8 ($CONH_2$), 169.0 ($HNOCCH_3$), 136.7 (C3), 135.3 (C7a), 128.0 (C3a), 120.7 (C6), 118.6 (C4), 118.2 (C5), 111.0 (C7), 106.8 (C3), 54.2 (t, J = 19, C8), 53.7 (Cα), 27.0 (Cβ), 22.6 ($NOCCH_3$),

HRMS-ESI(+): $[M+Na]^+$: found: 299.1337, calcd: 299.1225,

$[M+H]^+$: found: 277.1509 calcd 277.1405.

Model peptide synthesis

KFMAYH was synthesised using SPPS following a standard Fmoc-strategy with pre-attached Fmoc-L-His(Trt)-Wang resin (0.5 mmol/g) on an Activotec peptide synthesizer. An analogous procedure to that of KFWAWH peptide synthesis was followed (Chapter 4). The peptide was isolated as a white powder TFA salt (0.23 mmol, 47%), which was analysed by LC-UV to confirm >95% purity of the peptide - Fig. 5.32.

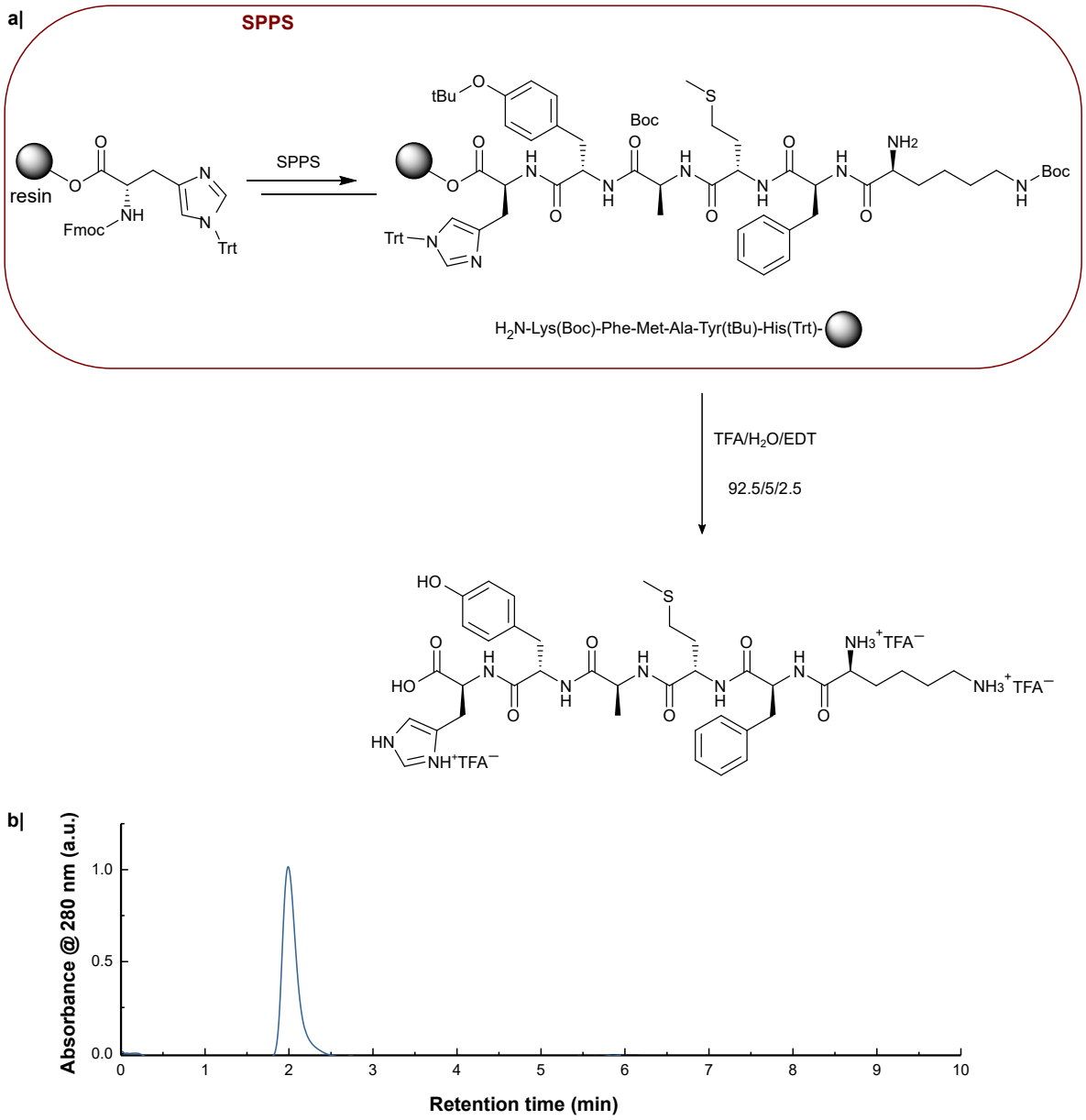


Fig. 5.32: **a|** Scheme of the KFMAYH model peptide synthesis. **b|** LC-UV chromatogram of the product (C18 Gemini NX (Phenomenex)), linear 10 mins gradient of MeCN in aqueous 0.1% formic acid.

The sequence was confirmed by MS/MS analysis (see elsewhere) and the peptide was used without further purification.

Model peptide reaction in the presence of redox active reagents

To peptide solution (9.2 μL, 10 mM in PBS), $PySO_2CF_2H$ (0.9 μL, 1 M in DMSO), additive (80 μL, 12 mM in PBS, in the case of urate the solid reagent was added to form a suspension) and $[Ru(bpy)_3]^{2+}$ (1.4 μL, 3.3 mM in PBS) were added. The reaction mixture was irradiated for 15 mins after which, an aliquot was diluted 10x with H_2O and analysed via LC-MS.

Model peptide reaction sequence - general procedure

To peptide solution (10 μL, 10 mM in PBS), $PySO_2CF_2H$ (2.5 μL, 25 eq, 1 M in DMSO), sodium ascorbate (1.5 μL, 25 eq, 1.65 M in PBS buffer), $[Ru(bpy)_3]^{2+}$ (1.5 μL, 5 mol%, 3.3 mM in PBS buffer) and PBS buffer (84.5 μL) were added. The reaction mixture was irradiated for 15 mins after which, an aliquot (2.5 μL) was diluted 10x with H_2O and analysed via LC-MS/MS. The rest of the reaction mixture was incubated at rt, after which it was analysed again by LC-MS/MS. Then, $NaBD_4$ (0.4 mg) was added to this solution and the reaction mixture was incubated for a further 24 h, when a final aliquot was analysed by LC-MS. An analogous procedure was used for the reaction with different peptides.

Difluoromethylation and tryptic digest of myoglobin

To a solution of myoglobin (100 μL, 1.8 mg/mL in PBS buffer), $PySO_2CF_2H$ (0.51 μL, 1 M in DMSO), sodium ascorbate (0.31 μL, 1.65 M in PBS) and $[Ru(bpy)_3]^{2+}$ (1.55 μL, 3.3 mM in PBS) were added. The reaction mixture was irradiated for 15 mins and after that, an aliquot (1 μL) was withdrawn and diluted with water (60 μL) for the LC-MS analysis. Note that initial conditions did not use the sacrificial amine. To

the remaining reaction mixture, water (41 μL) was added before the buffer exchange with a MiniTrap G25 (into 50 mM NH_4HCO_3). To perform the digest, trypsin was added (3.5 μL, 1:25 w/w, 2 mg/mL in 50 mM acetic acid) and the reaction mixture was incubated overnight at 37 °C. After that, the tryptic peptides were bound on a C18 ZipTip, washed, eluted according to manufacturer's instructions and analysed via LC-MS/MS. LC separation with a C18 BEH Agilent column was achieved using the following gradient: 95% A at (0 - 2) min, gradient to 60% B at (2.1 - 30) min, 95% B at (30.1– 33) min, 95% A at (33.1– 35) min with a flow rate of 0.3 mL/min.

Difluoromethylation and tryptic digest of lysozyme

To a solution of lysozyme (100 μL, 1.5 mg/mL in PBS/Lys buffer - with 100 mM lysine), $PySO_2CF_2H$ (1 μL, 1 M in DMSO), sodium ascorbate (0.6 μL, 1.65 M in PBS/Lys) and $[Ru(bpy)_3]^{2+}$ (1.55 μL, 3.3 mM in PBS/Lys) were added. The reaction mixture was irradiated for 15 mins and after that an aliquot (1 μL) was withdrawn and diluted with water (50 μL) for the LC-MS analysis. To the remaining reaction mixture, water (41 μL) was added before the buffer exchange with a MiniTrap G25 (into 50 mM NH_4HCO_3). To this solution, DTT (1.4 μL, 1 M in 50 mM NH_4HCO_3) was added and the reaction mixture was incubated for 40 mins at 56 °C. Next, iodoacetamide (30 μL, 110 mM in 50 mM NH_4HCO_3) was added and the protein was alkylated for 30 mins in the dark at rt. Then, DTT was added (1.7 μL, 1 M in 50 mM NH_4HCO_3) to quench the unreacted iodoacetamide. Finally, trypsin was added (3 μL, 1:25 w/w, 2 mg/mL in 50 mM acetic acid) and the reaction mixture was incubated overnight at 37 °C. After that, the tryptic peptides were bound on a

C18 ZipTip, washed, eluted and analysed via LC-MS/MS as per above conditions.

Protein difluoromethylation/hydrolysis/labelling sequence general procedure

1. To a solution of a protein (484 μL, 105 μM, in PBS buffer containing 100 mM Lys), $PySO_2CF_2H$ (5.1 μL, 1 M in DMSO), sodium ascorbate (3.1 μL, 1.65 M in PBS/Lys) and $[Ru(bpy)_3]^{2+}$ (7.7 μL, 3.3 mM in PBS/Lys) were added. The reaction mixture was irradiated for 15 mins and after that an aliquot (1 μL) was withdrawn and diluted with water (60 μL) for LC-MS analysis. In the case of lactalbumin, instead of water the aliquot for MS analysis was diluted with DTT solution (5 mM) to improve the protein ionisation. Note that this procedure can be scaled down in volumes proportionally.

2. The remaining reaction mixture was passed through a MiniTrap G25 (equilibrated in 8 M guanidine hydrochloride PBS pH 7.3) and eluted with the same buffer (1 mL).

   (a) For α-nucleophile labelling: an α-nucleophile (1 M in DMSO) was added to the protein solution so that its final concentration was 10 mM and the reaction mixture was incubated at 37 °C overnight (c. 14 h).

   (b) For reductive amination: an amine (1 M in DMSO or PBS buffer) was added to the protein solution so that final concentration was 20 mM. The reaction mixture was incubated at 37 °C for 3 h before $NaBH_3CN$ (5 M in PBS) was added (50 mM final concentration). The mixture was then further incubated overnight at 37 °C.

Protein difluoromethylation/hydrolysis/borodeuteride reduction sequence

1. Difluoromethylation following the above general procedure.

2. The remaining reaction mixture was passed through a MiniTrap G25 (equilibrated in 100 mM NH_4OAc, 8 M guanidine hydrochloride, pH 6.0) and eluted with the same buffer (1 mL). After an overnight incubation at 37 °C, an aliquot (75 μL) of the reaction mixture was treated with freshly prepared $NaBD_4$ solution (25 μL, 3.4 mg/mL in ice cold water) and the reaction mixture was incubated at rt for 6 h before LC-MS analysis.

Anti-biotin Western Blot/Dot blot

The gel after SDS-PAGE analysis of the modified proteins (10 μg) was electrophoretically blotted onto a membrane (iBlot Gel Transfer Device, Thermo Fisher) or for dot blots, the protein biotinylation crude reaction mixtures were spotted onto a membrane (1 μL, 50 μM the crude reaction mixture). The membrane was blocked with 5% bovine serum albumin (BSA) in TBST (20 mM TRIS, pH 7.5, 150 mM NaCl, 0.1% Tween 20) for 1 h and then washed with TBST (3 x 20 mL) with 5 mins incubation periods at rt. After that, a solution the anti-biotin antibody (*A7064 Sigma*, 2 μL in 20 mL of 1% BSA in TBST) was poured over the membrane, which was then incubated at 4 °C overnight. Finally, the membrane was washed with TBST (20 mL, 5 mins) and TBS (30 mL, as TBST but without Tween 20). Visualisation was achieved with the BCIP/NBT liquid substrate system (3 mL, *72091 Sigma-Aldrich*) and images of developing bands were taken periodically in 15 min periods.

5.7 Appendix

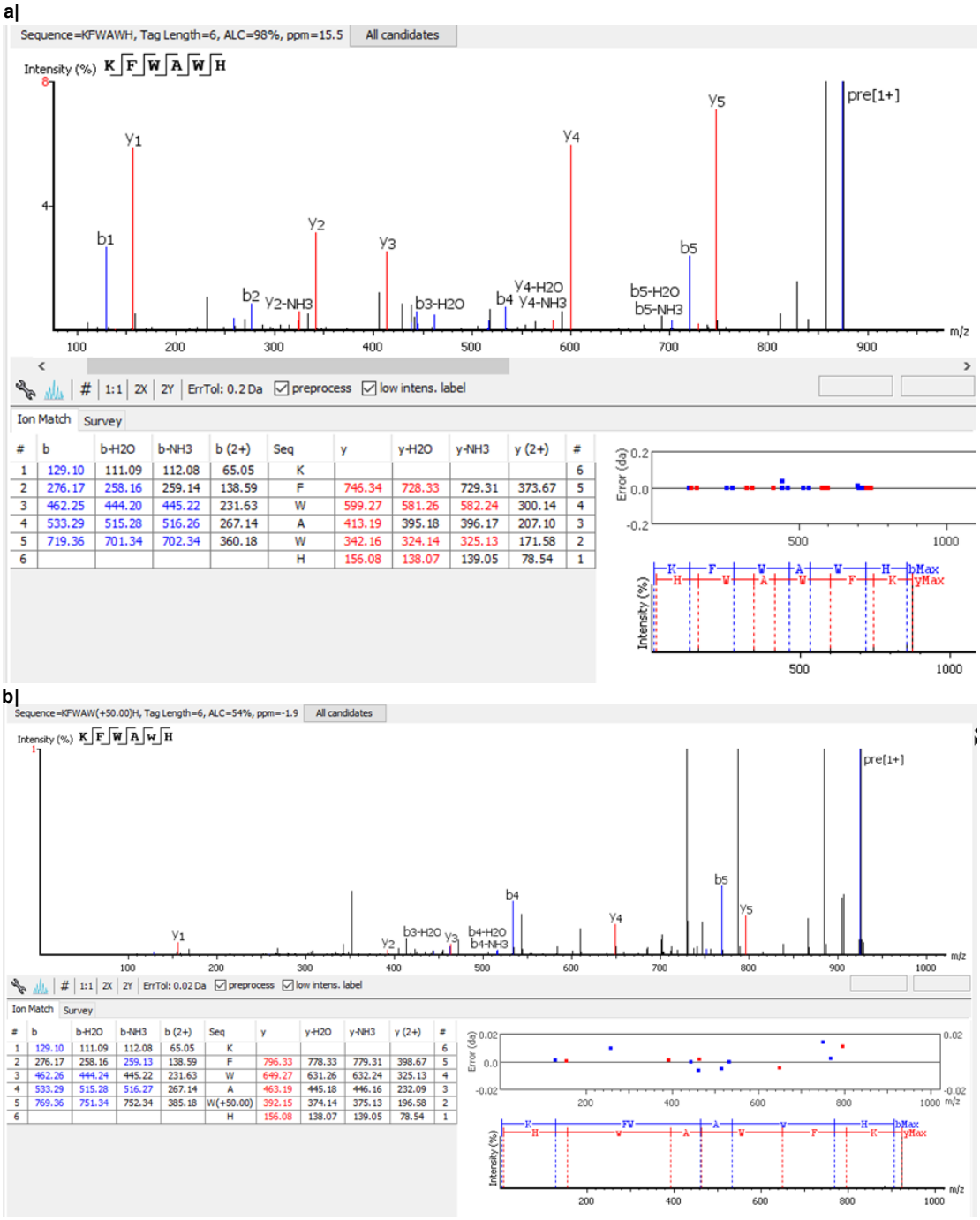


Fig. 5.33: Tandem mass spectra of different m/z precursor ions found in the difluoromethylation reaction mixture of KFWAWH peptide. **a**| $m/z = 874$ ion (starting material) demonstrating original fragmentation pattern with ions assignment, **b**| $m/z = 924$ ion (+50) revealing Trp5 difluoromethylation at RT = 5.5 min.

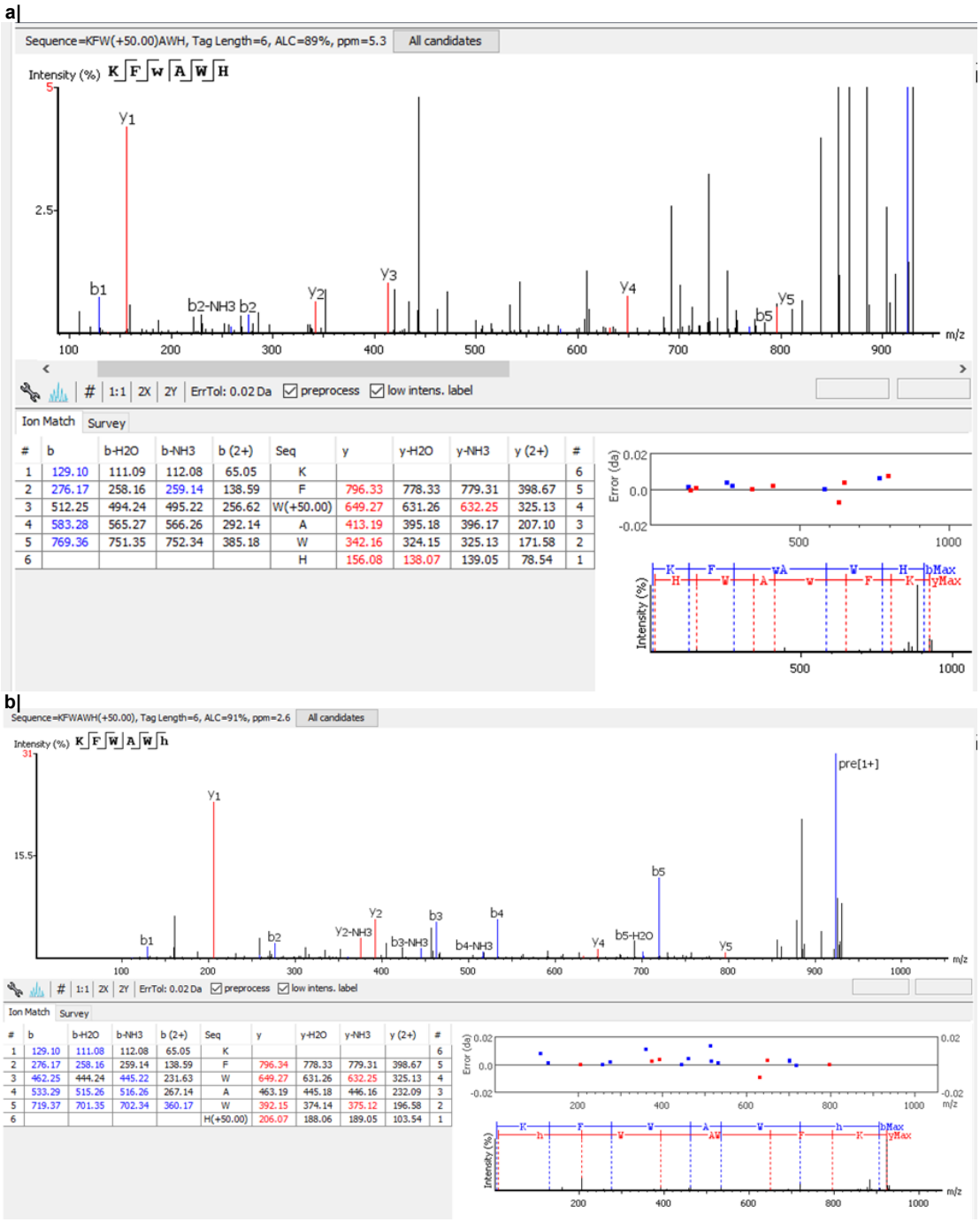


Fig. 5.34: Tandem mass spectra of different m/z precursor ions found in the difluoromethylation reaction mixture of KFWAWH peptide. **a**| $m/z = 924$ ion (+50) revealing Trp3 difluoromethylation at RT = 5.5 min, **b**| $m/z = 924$ ion (+50) revealing His6 difluoromethylation at RT = 5.75 min.

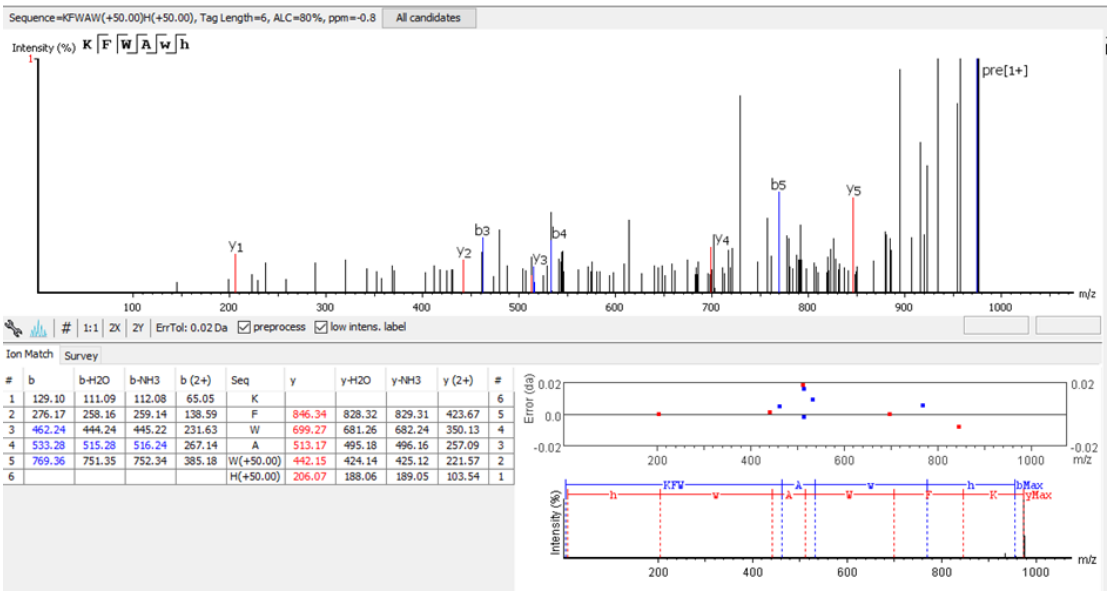


Fig. 5.35: Tandem mass spectra of m/z = 974 precursor ion found (+100) in the difluoromethylation reaction mixture of KFWAWH corresponding to Trp5 and His6 double difluoromethylated product at RT = 5.9 min.

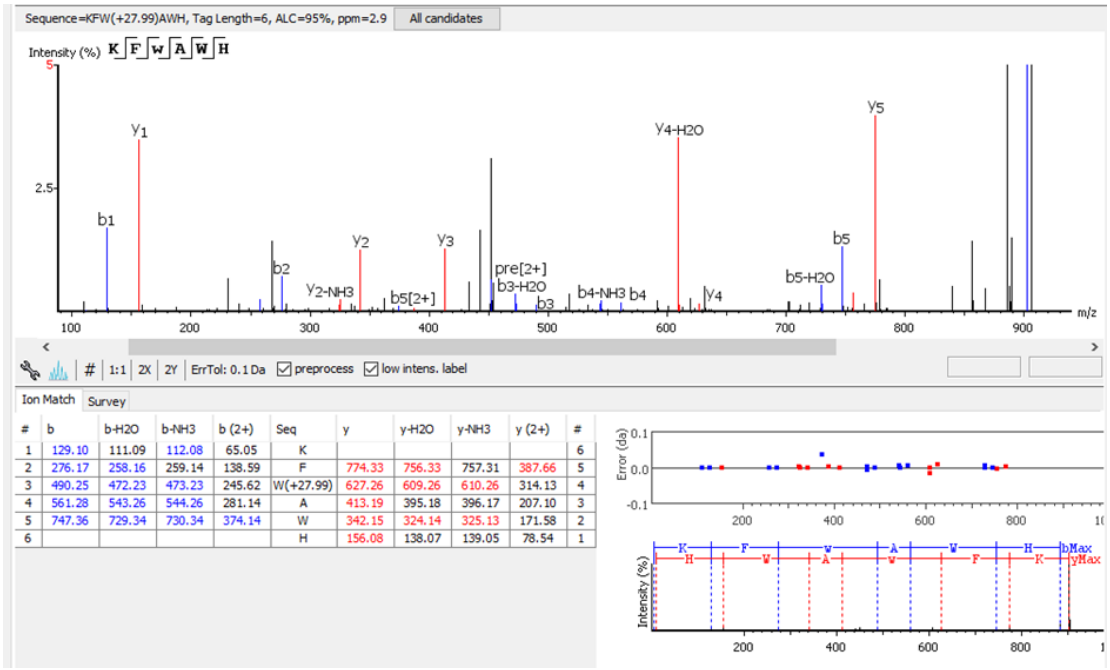
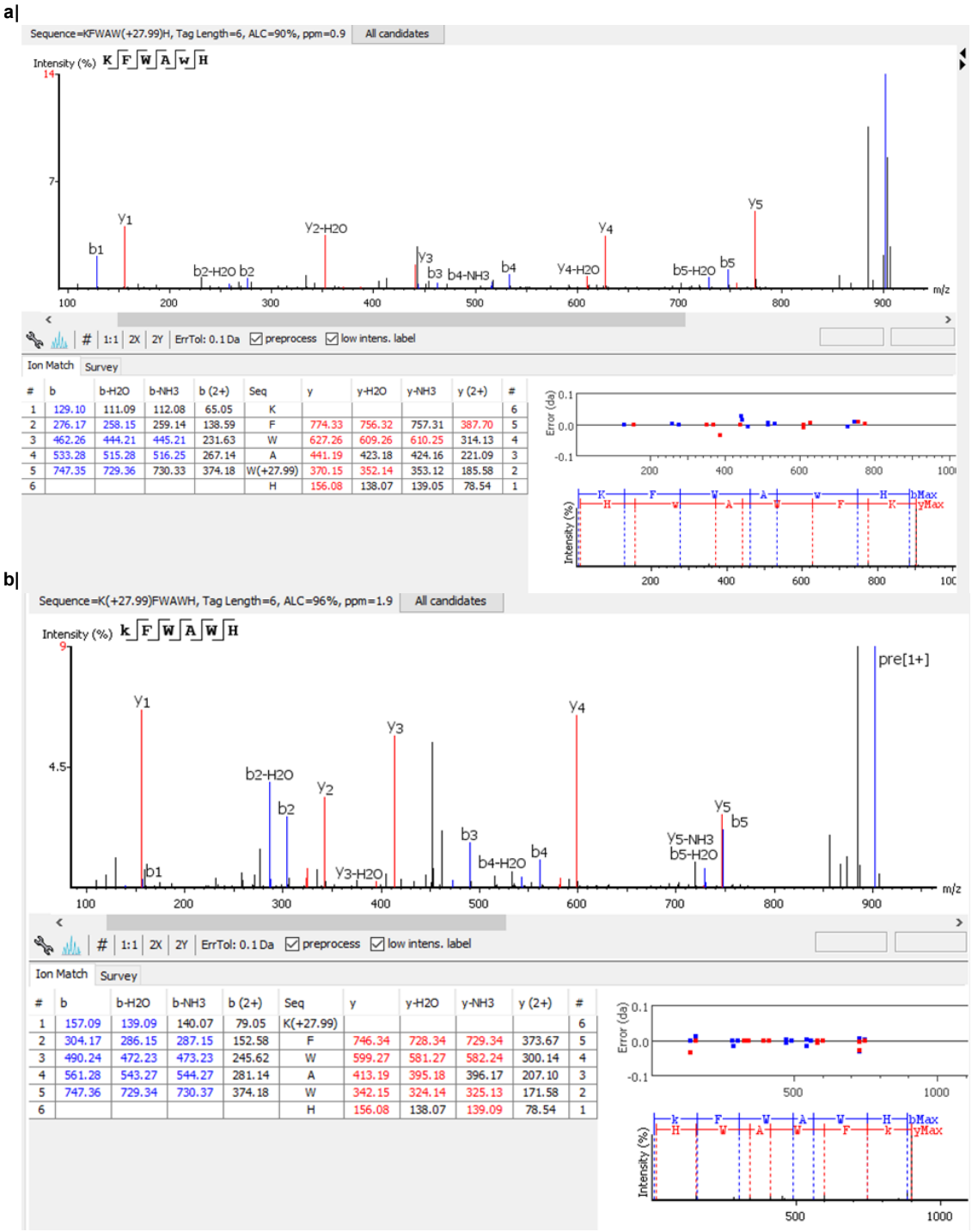


Fig. 5.36: Tandem mass spectra of m/z = 902 precursor ion (+28) found in the difluoromethylation reaction mixture of KFWAWH peptide corresponding to formylated Trp3 product at RT = 5.2 min.



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.09	112.08	65.05	K					6
2	276.17	258.15	259.14	138.59	F	774.33	756.32	757.31	387.70	5
3	462.26	444.21	445.21	231.63	W	627.26	609.26	610.25	314.13	4
4	533.28	515.28	516.25	267.14	A	441.19	423.18	424.16	221.09	3
5	747.35	729.36	730.33	374.18	W(+27.99)	370.15	352.14	353.12	185.58	2
6					H	156.08	138.07	139.05	78.54	1

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	157.09	139.09	140.07	79.05	K(+27.99)					6
2	304.17	286.15	287.15	152.58	F	746.34	728.34	729.34	373.67	5
3	490.24	472.23	473.23	245.62	W	599.27	581.27	582.24	300.14	4
4	561.28	543.27	544.27	281.14	A	413.19	395.18	396.17	207.10	3
5	747.36	729.34	730.37	374.18	W	342.15	324.14	325.13	171.58	2
6					H	156.08	138.07	139.09	78.54	1

Fig. 5.37: Tandem mass spectra of different m/z precursor ions found in the difluoromethylation reaction mixture of KFWAWH peptide. **a**| m/z = 902 ion (+28) revealing Trp5 formylation at RT = 5.3 min, **b**| m/z = 902 ion (+28) revealing Lys1 formylation at RT = 5.6 min.

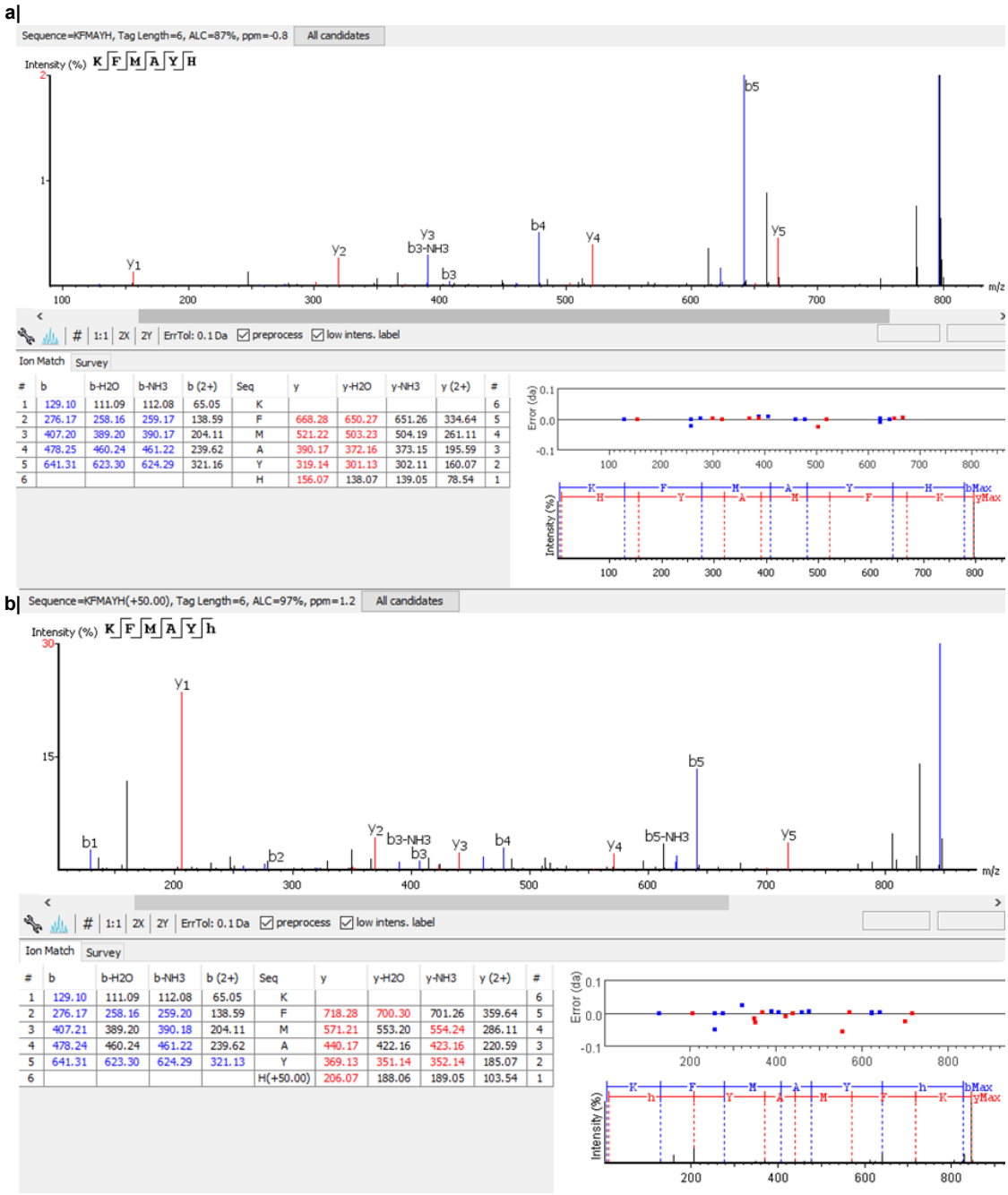


Fig. 5.38: Tandem mass spectra of different m/z precursor ions found in the difluoromethylation reaction mixture of KFMAYH peptide. **a**| $m/z = 796$ ion (starting material) demonstrating original fragmentation pattern with ions assignment, **b**| $m/z = 846$ ion (+50) revealing His6 difluoromethylation at RT = 3.2 min.

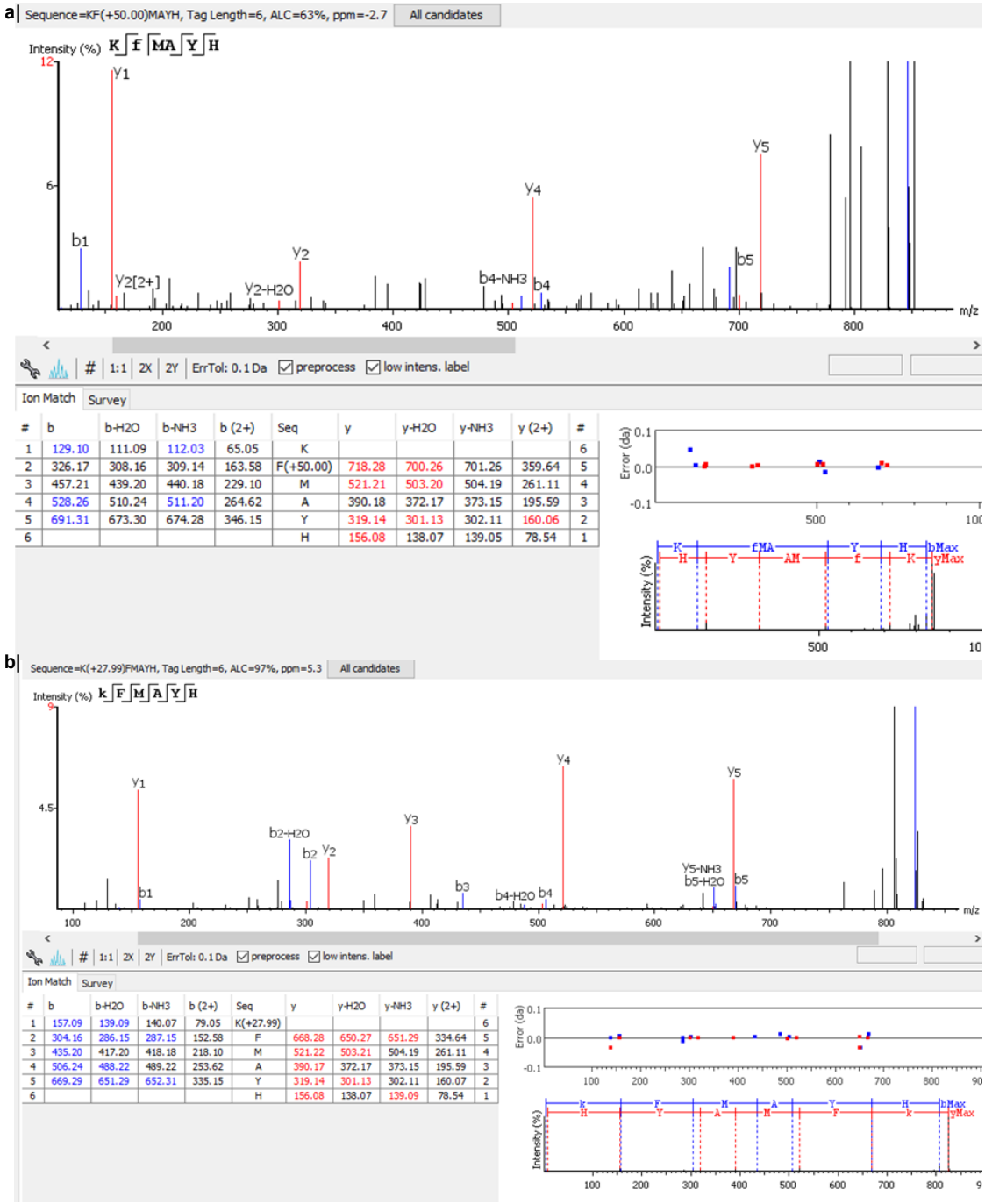
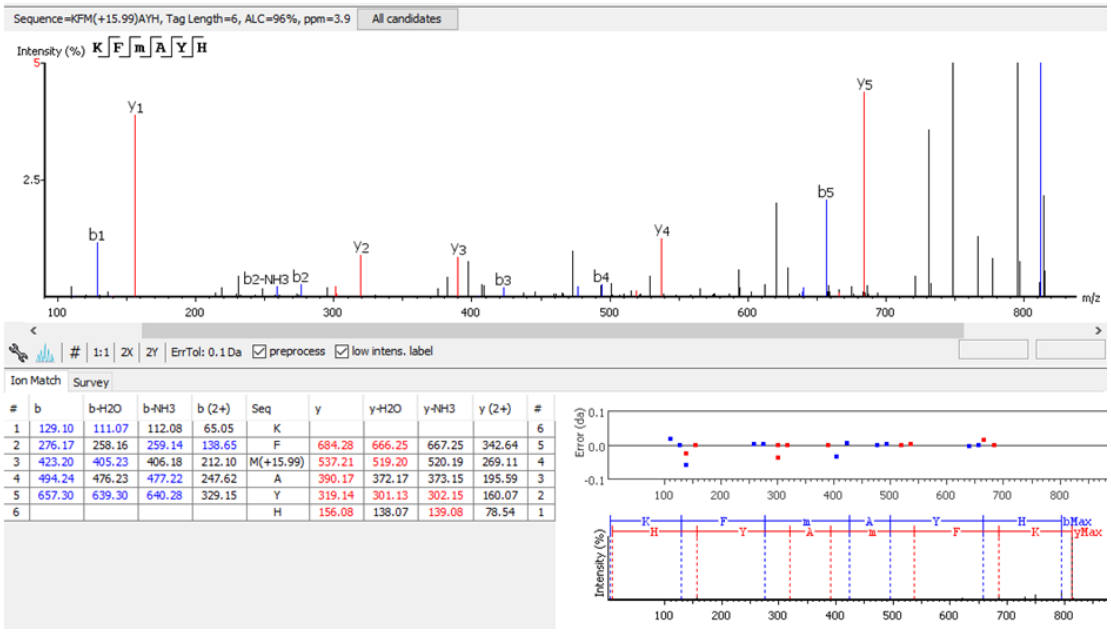


Fig. 5.39: Tandem mass spectra of different m/z precursor ions found in the difluoromethylation reaction mixture of KFMAYH peptide. **a**| m/z = 846 ion (+50) revealing Phe2 difluoromethylation at RT = 2.9 min, **b**| m/z = 824 ion (+28) revealing Lys1 formylation at RT = 3.25 min.



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.07	112.08	65.05	K					6
2	276.17	258.16	259.14	138.65	F	684.28	666.25	667.25	342.64	5
3	423.20	405.23	406.18	212.10	M(+15.99)	537.21	519.20	520.19	269.11	4
4	494.24	476.23	477.22	247.62	A	390.17	372.17	373.15	195.59	3
5	657.30	639.30	640.28	329.15	Y	319.14	301.13	302.15	160.07	2
6					H	156.08	138.07	139.08	78.54	1

Fig. 5.40: Tandem mass spectra of m/z = 812 precursor ion found (+16) in the difluoromethylation reaction mixture of KFMAYH peptide corresponding to the Met2 oxidised product at RT = 2.3 min.

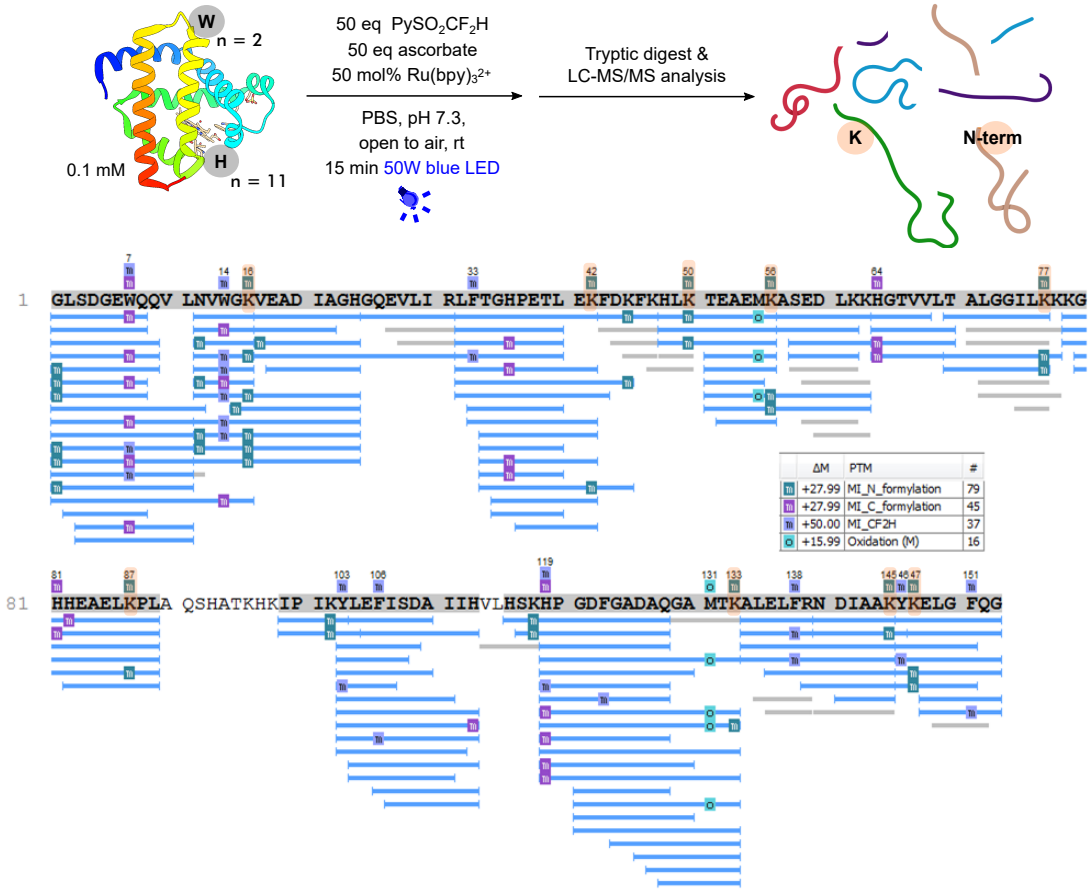


Fig. 5.41: LC-MS/MS analysis of the tryptic peptides for myoglobin difluoromethylated under initial conditions (no competitive amine). Good peptide map coverage (FDR<0.5%, PEAKS software) was obtained, expected modifications were colour coded as per legend, revealing prominent N-formylation (N-terminus, Lys; 79 instances within 9 confident sites) among the expected C-formylation and residual difluoromethylation.

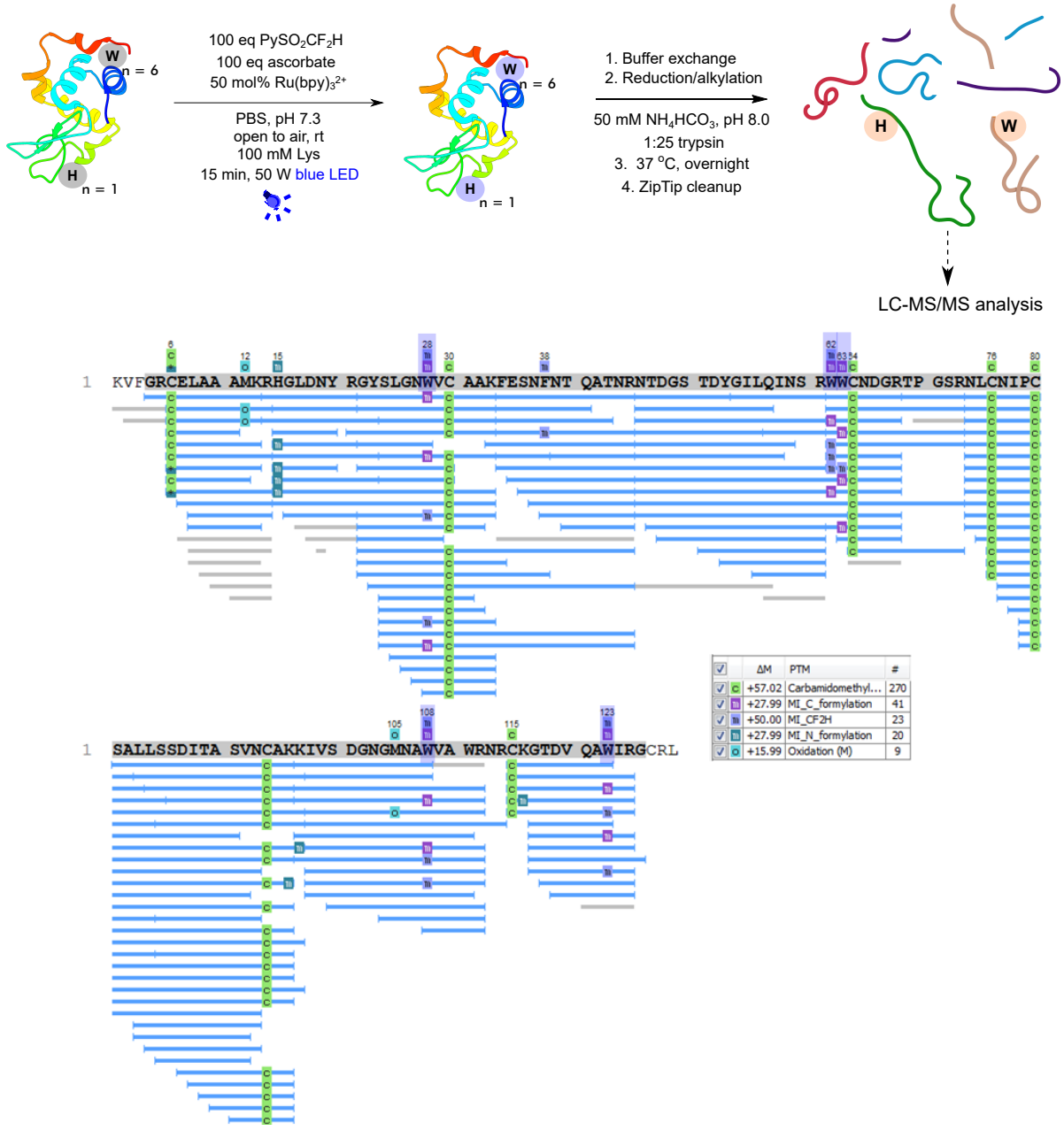


Fig. 5.42: LC-MS/MS analysis of the tryptic peptides for difluoromethylated lysozyme. Before the digest, the protein disulfide bonds were reduced and free Cys alkylated with iodoacetamide (bright green modification). Good peptide map coverage (FDR<0.5%, PEAKS software) was obtained, expected modifications were colour coded as per legend, revealing the expected Trp C-formylation as the dominant modification (41 instances within 5 confident residues). Note that no CF_2H modifications were encountered for Cys residues.

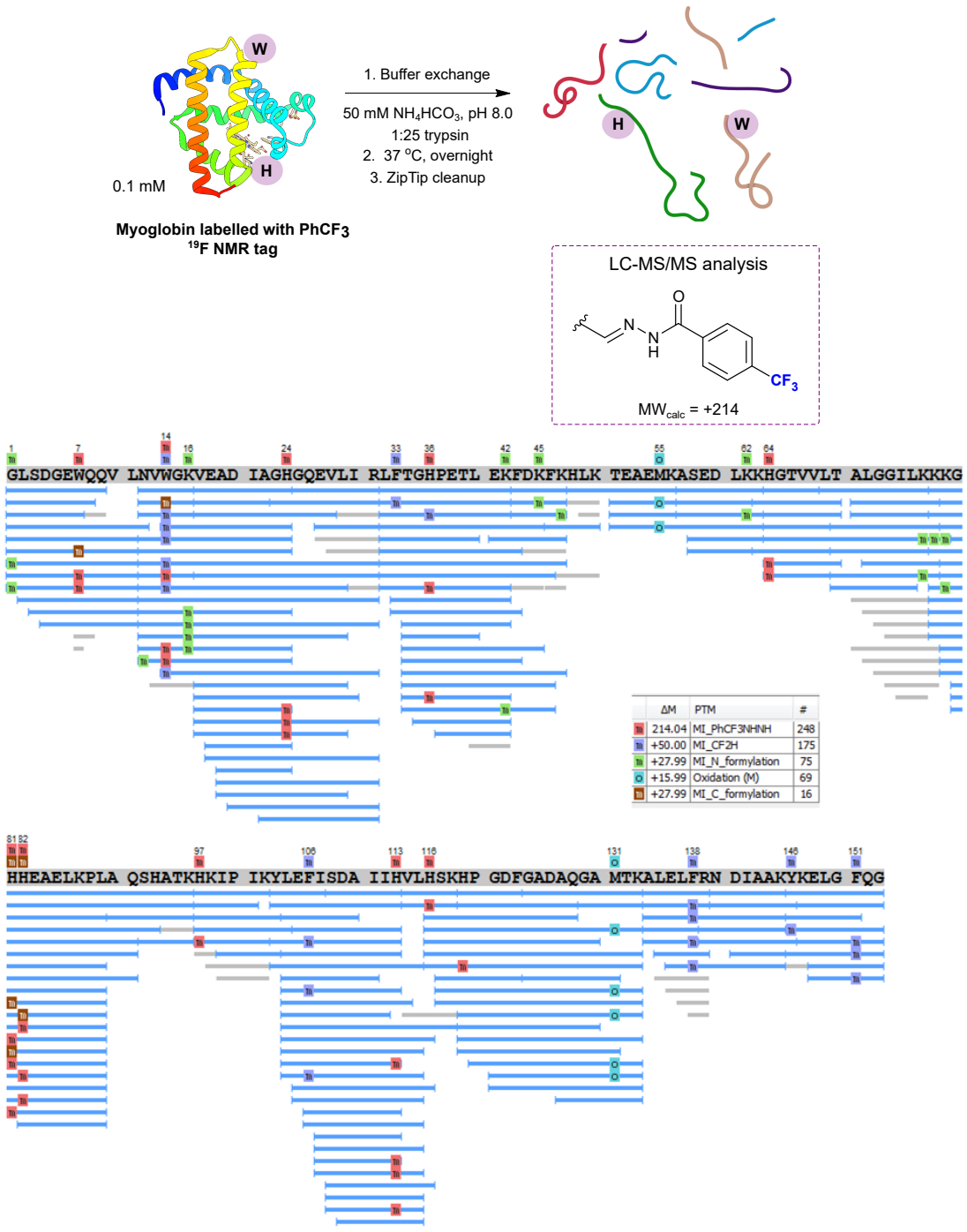
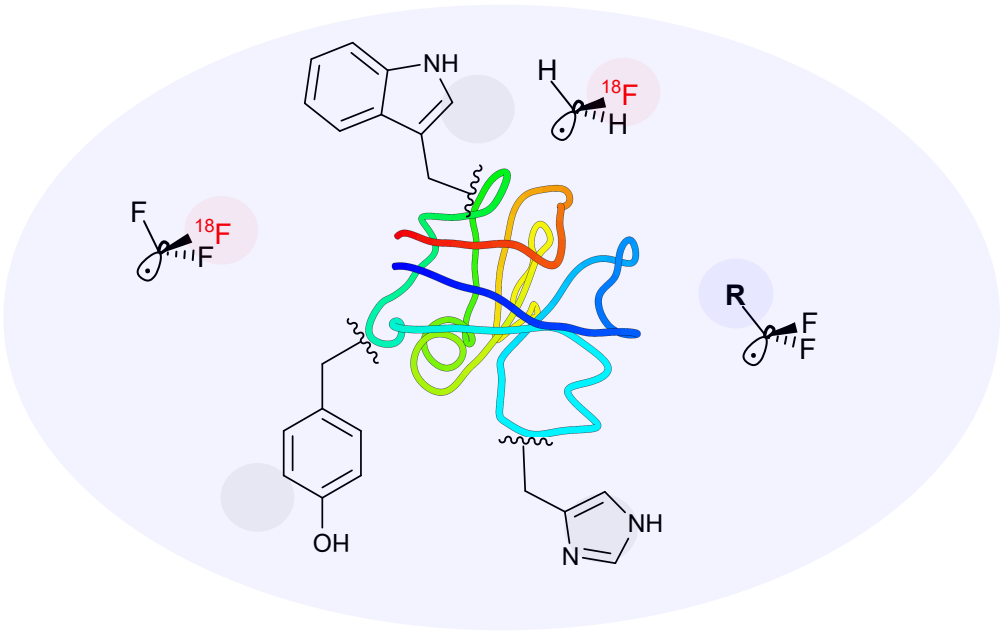


Fig. 5.43: LC-MS/MS analysis of the tryptic peptides for myoglobin labelled with ^{19}F NMR hydrazone tag. Good peptide map coverage (FDR<0.5%, PEAKS software) was obtained, expected modifications were colour coded as per legend, revealing expected hydrazone extension (red squares) as the dominant modification across expected residues. Label free quantification (PEAKS) was used to evaluate the relative extent of each residue modification.

Chapter 6

Outlook and perspectives for radical fluoroalkylations of proteins



Abstract

Chapter 6 outlines the perspectives for radical fluoroalkylations of proteins and evaluates chemistry developed throughout this thesis. The unsolved problem of protein radiotrifluoromethylation is considered, and preliminary results showing the promise of photoredox chemistry are disclosed. Furthermore, the light driven radical generation platform is shown to be amenable for the screening of a variety of fluoroalkyl radicals, using different classes of radical precursors: benzathiazole or pyridyl sulfones or alkyl difluorobromides. As such a clear direction is shown, opening the doors for installation and study of different fluoroalkyl chains.

6.1 Perspectives for protein fluoroalkylation

The work presented in this thesis has aimed at pushing the boundaries of protein chemistry by investigating non-traditional modes of single electron reactivity to modify biomolecules. Acknowledging the great progress that has been made in the last decades in terms of bioconjugation and precise techniques for editing biomolecules, this work focused on the investigation of unexplored reactivity of fluoroalkyl radicals with native protein residues. This has resulted in several new strategies being validated with a number of Trp selective fluoroalkylation protocols proposed. This chapter intents to discuss future perspectives and offer insights into the possible developments to the field that could be based off this work.

Chapter 2 discussed reactivity and selectivity of the trifluoromethyl radical, generated oxidatively from a sulfinate precursor, in reactions with amino acids, peptides and proteins. Initial experiments revealed an interesting feature of this chemistry - preferential targeting of one structural motif, in this case Trp, over others. This allowed for the use of the CF_3 group as a selective and minimally sized ^{19}F NMR tag, as the specific application of this methodology. With several examples and proof-of-concept NMR experiments the project was successfully concluded. Although having stirred interest in the scientific community (34 citations and 8111 articles views according to ACS website, August 2020)[232], the method is yet to be applied to facilitate specific chemical biology experiments, e.g. ligand discovery or structural investigation of a Trp residue role. However, with the widespread awareness of the potential and a recent resurgence of interest in protein observed ^{19}F NMR[392], it is predicted that

this will occur in the coming years.

6.1.1 Benzothiazole sulfones for radiofluorination

Building on this universal methodology, attempts were made to extend the protocol to ^{18}F-trifluoromethylation of proteins - the topic of Chapter 3. However, difficulties in the radiosynthesis of the radical precursor, as well as a seemingly inefficient radical initiation mode prevented the fruitful realisation of the idea. Although a variant of the method has been efficiently applied to the modification of peptides[320], suitable conditions that could be used for protein substrates were not found nor were reported. However, taking inspiration from the developments described in Chapters 4 and 5, regarding the alternative radical generation system, it was proposed that protein radiotrifluoromethylation using radical chemistry might be realised using heteroaryl sulfones as more efficient radical precursors.

The hypothesis was deemed worthy of exploration, as a suitably ^{18}F-labelled precursor was prepared as a part of the developed radiosynthesis - namely **BzSOOCF$_3$** (see Fig. 3.7). This fluoroalkyl compound belongs to the same class as pyridyl sulfones - **PySO$_2$CF$_2$R**, extensively used in Chapters 4 and 5. Moreover, its suitability for protein chemistry might be even higher, as it has a lower reduction potential, therefore, easily accessible by mild photocatalysts[177]. Unfortunately, due to the operational reasons, it was not possible to test this compound in a radiochemistry settings. However, the reaction was tested in the ^{19}F feasibility studies. A straightforward application of conditions developed for CF$_2$H-ylation (see Fig. 5.14) with the **BzSOOCF$_3$** as the radical precursor and myoglobin as the substrate, resulted in

the formation of trifluoromethylated proteins products as observed by LC-MS (data not shown). Next, the reaction was repeated under conditions resembling those used during radiofluorination, with the radical precursor as the limiting reagent and an excess of other substrates - Fig. 6.1.

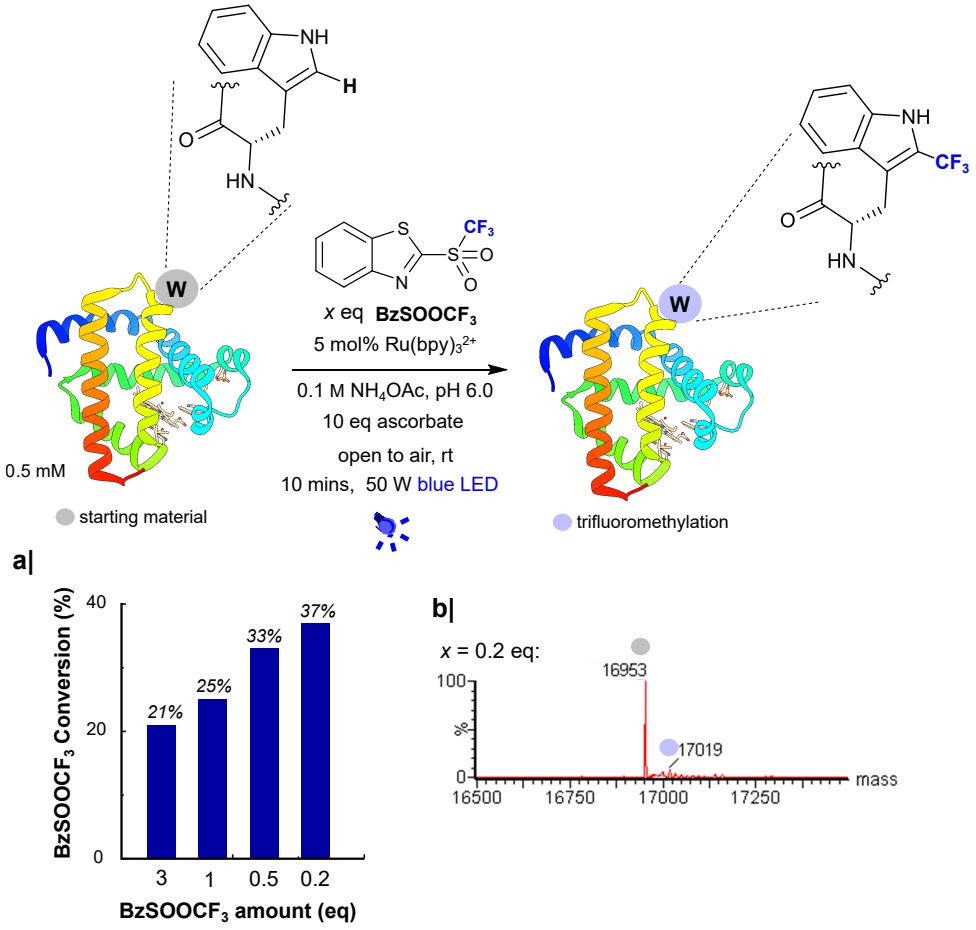


Fig. 6.1: Photoredox mediated trifluoromethylation of myoglobin with the benzothiazole sulfone radical precursor. Stoichiometry of reagents in the scheme given in relation to the protein. **a**| LC-MS quantification of the CF_3 protein product was used to calculate relative conversions of the radical precursor substrate. **b**| Example of mass spectrum of the reaction mixture where substoichiometric amounts of **BzSOOCF$_3$** were used.

The experiment revealed superior reactivity of this precursor under photoredox activation, as compared to the sulfinate/oxidant system, which required superstoichiometric

amounts to observe modified protein (Fig. 2.24). Here, use of as little as 0.2 eq was enough to generate a detectable conversion - Fig. 6.1b. Interestingly, reducing the amount of the precursor, resulted in higher efficiency of the reaction with respect to the radical precursor, suggesting that this protocol might be indeed suitable for ^{18}F-labelling - Fig. 6.1a.

It is important to note that this class of substrates had been already used for radio-*difluoromethylation* of small molecule heterocycles under a photoredox activation manifold[321]. On the other hand, the challenge of the efficient [^{18}F]-CF_3 group radiosynthesis still persists. However, considering the fact that the loss of fluorine might be inhibited for folded *difluoromethylated* proteins, using the same [^{18}F]-CF_2H heteroaryl sulfone radical precursors may also lead to applications in protein ^{18}F-radiolabelling - Fig. 6.2a. Advantageously, the radiosynthesis of such radical precursors i.e. **[^{18}F]-BzSOOCF$_2$H** has been achieved with high molar activity, making such a transformation a more desirable to pursue[393]. Should the loss of fluorine from the modified protein be significant enough to cause an increase in the background activity, further studies on CF_2R radicals possessing R groups stabilising the fluoroalkylated heteroaromatic products might be necessary. Finally, positioning of a specific radical acceptor, which modification would lead to a stable product (e.g. dehydroalanine)[53], can be considered - Fig. 6.2b. Recently developed efficient photoredox difluoromethylation protocol could be envisioned for radiodifluoromethylation of Dha-containing proteins[394]. Such approaches are likely to be conducive to achieving protein radiofluorination using the radical chemistry explored in this thesis.

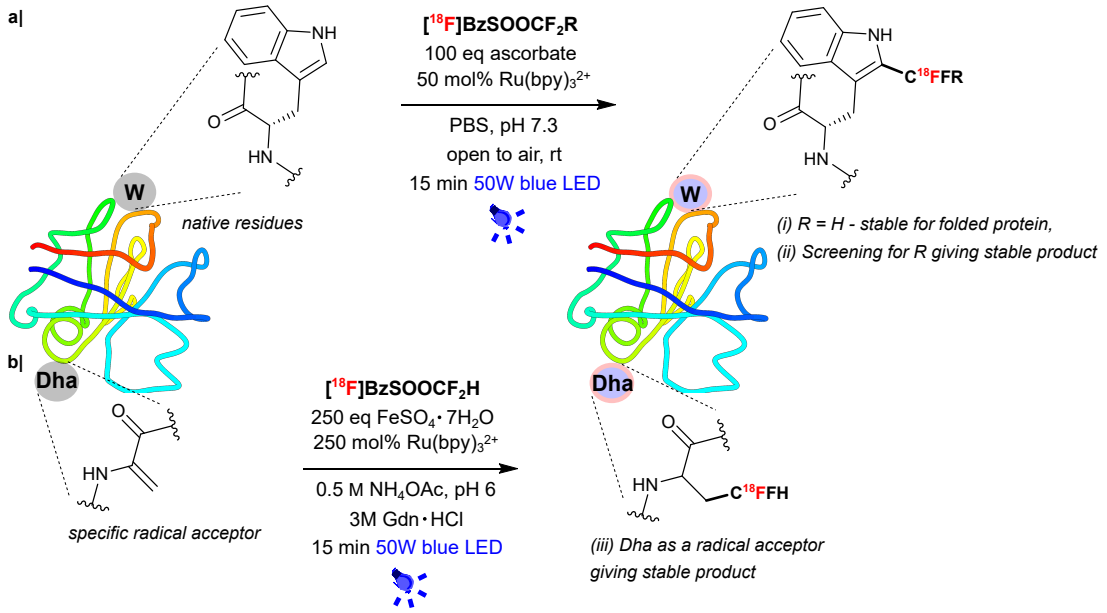


Fig. 6.2: Potential approaches for alternative photoredox radiofluorination of proteins. **a**| [^{18}**F**]-**BzSOOCF$_2$R**/ascorbate system for native residues radiolabelling. **b**| [^{18}**F**]-**BzSOOCF$_2$H**/Fe^{2+} system for the engineered Dha residue radiolabelling[394].

6.1.2 Pyridyl sulfones as fluoroalkyl radical precursors

Part of the work in this thesis was focused on investigating the general reactivity of fluoroalkyl radicals with native protein residues. Side reactivity of the sulfinate radical generation system was attributed to the instability of the difluoroalkyl sulfinates in aqueous oxidative conditions - Chapter 4. Although curiously, while these compounds were found to be suitable reagents for N-acylation of proteins, this is not predicted to have a significant impact in the wider community, as such a modification is an essentially solved problem in protein chemistry. To continue the investigation, a different platform for the biocompatible generation of fluoroalkyl radicals was established, based on the stable radical precursors - namely pyridyl sulfones. Study of the model radical bearing an alkyl substituent, $\cdot CF_2CH_3$, revealed its poor reactivity and

selectivity patterns and the project was not further elaborated. By the same token, mounting evidence from the literature supported these results as discussed in Chapter 4. Therefore, further investigations were performed on the CF_2H radical, which although different in nature from its trifluoromethyl counterpart, was found to follow similar reactivity patterns (preferential reaction with Trp) under the proposed conditions. However, of most importance, the difluoromethylated protein product was found to be hydrolytically unstable - leading to the formation of the corresponding acyl derivative. This was exploited to create a new method for the introduction of an aldehyde tag into proteins, and the manuscript describing this work was submitted for publication.

It remains an interesting and open question, as to what is the influence of the R group in CF_2R radicals, with respect to their reactivity, selectivity and product stability in the context of protein chemistry. From another perspective, the feasible products also have to find applications in chemical biology experiments. Certain evidence can be gathered from recent reports, for example, an immediate extension of the chain with the additional fluorine atoms ($R = CF_2R_1$) resulted in an efficient protein reaction, leading to stable products with a seemingly enhanced Trp selectivity[123], establishing a new bioconjugation strategy.

Although outside the immediate scope of this thesis, some experiments were performed to demonstrate that the developed photoredox radical generation platform might be used for a general fluoroalkyl radicals screening study. One of the main benefits of the pyridyl sulfone system is that it is well-poised for the generation of a range of different fluoroalkyl derivatives (radical precursors), as demonstrated exten-

sively in the literature[177,395,396]. Here, a new reagent was synthesised to test whether the CF_2Ph radical could be used for the protein modification, yielding upon the hydrolysis a potentially photoactivable aryl ketone system that could be explored for cross-linking applications[166] - Fig. 6.3.

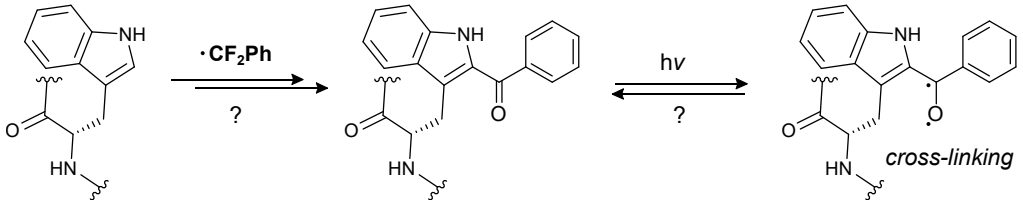
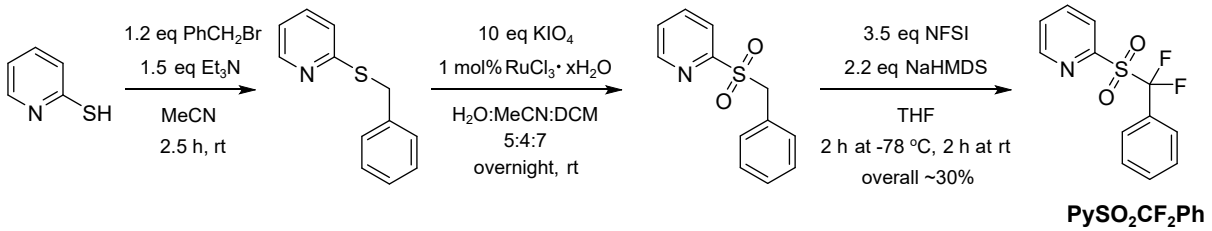


Fig. 6.3: A potential aryl ketone product, derived from hydrolysed fluoroalkylated product, might be useful in light induced protein cross-linking.

The required compound, $\mathbf{PySO_2CF_2Ph}$, was easily synthesised following a multistep route - Sch. 6.1.



Sch. 6.1: Synthesis of the $\mathbf{PySO_2CF_2Ph}$ radical precursor via sequence of alkylation, oxidation and fluorination.

When reacted with the model protein myoglobin under standard or fortified difluoromethylation conditions (see Fig. 5.14), no products could be observed by LC-MS (data not shown). Similarly, essentially no products (<1%) could be detected for the model peptide KFWAWH, suggesting that difluorobenzylation cannot be readily performed for native protein residues, seemingly due to the insufficient reactivity of this species in the proposed system - Fig. 6.4. Unfortunately, this reaction could not

be forced any further by increasing reagent concentration due to its poor solubility or by an increase of irradiation time (up to 1 h).

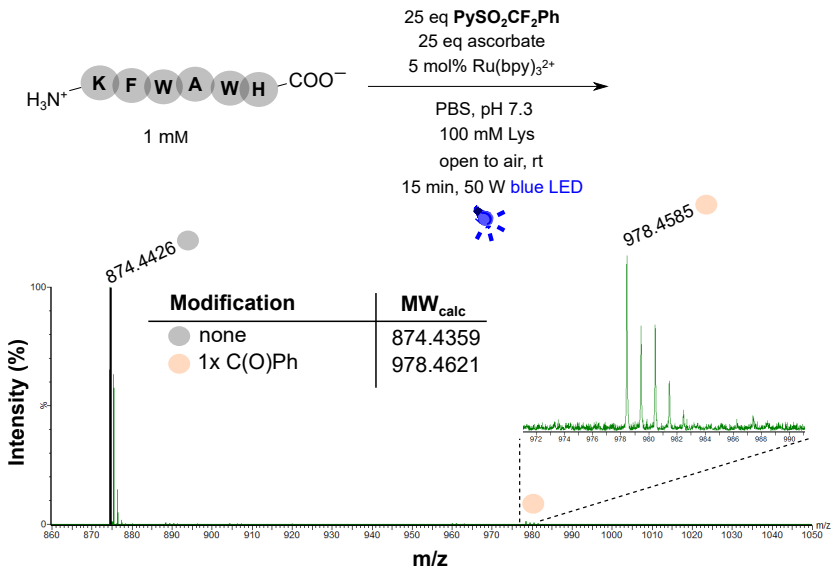


Fig. 6.4: Difluorobenzylation of the model peptide KFWAWH with the pyridyl sulfone precursor under photoredox activation. Combined mass spectrum of the reaction mixture revealed that traces of the hydrolysed product were formed.

Since aryl ketones of different structures (e.g. acetophenone[397]) have also been shown to be reactive in photo cross-linking assay, the idea might be realised upon screening different fluoroalkyl radicals and characterisation of the hydrolysed protein products in the view of cross linking reactivity.

6.1.3 Alkyl difluorobromides as fluoroalkyl radical precursors

In order to facilitate rapid screening of radicals, it was wondered whether a different class of precursors (with better commercial availability), could be activated using similar photoredox conditions in the reductive manifold. Perfluoroalkyl halides have been shown to be reactive following this chemistry, efficiently generating a range

of radical species[181]. To evaluate their potential, a simple precursor $BrCF_2CO_2Na$, potentially giving access to the α-ketoacid motif useful in ligation methodologies[398], was tested in a reaction with the model peptide - Fig. 6.5 and Fig. 6.6.

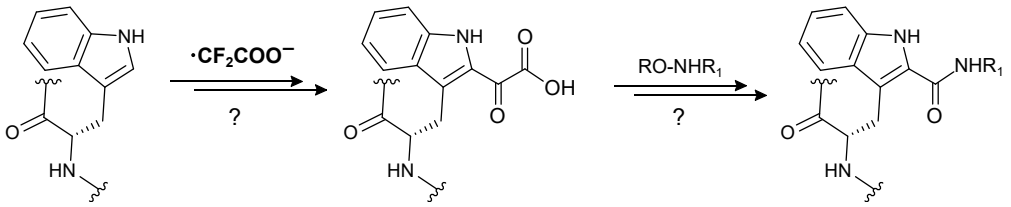


Fig. 6.5: A potential α-ketoacid product, derived from the hydrolysed fluoroalkylated product, might be useful in KAHA ligation methodologies.

Under the developed photoredox conditions (see Chapters 4 and 5), the formation of the expected radical addition and hydrolysis product was observed by LC-MS. To prove that the observed hydrolysis products constituted C-acyl derivatives, the previously developed borodeuteride assay was carried out (see Chapter 4.4.1). Gratifyingly, it revealed full conversion to the reduced products. Interestingly, the peaks corresponding to difluoroalkylated products remained present in the spectrum after 24 h of incubation, indicating incomplete hydrolysis and suggesting their increased stability in aqueous conditions. This experiment aimed to be a proof-of-concept, and full characteristics (e.g. such as a detailed selectivity pattern) of the $CF_2CO_2^-$ radical were not elucidated.

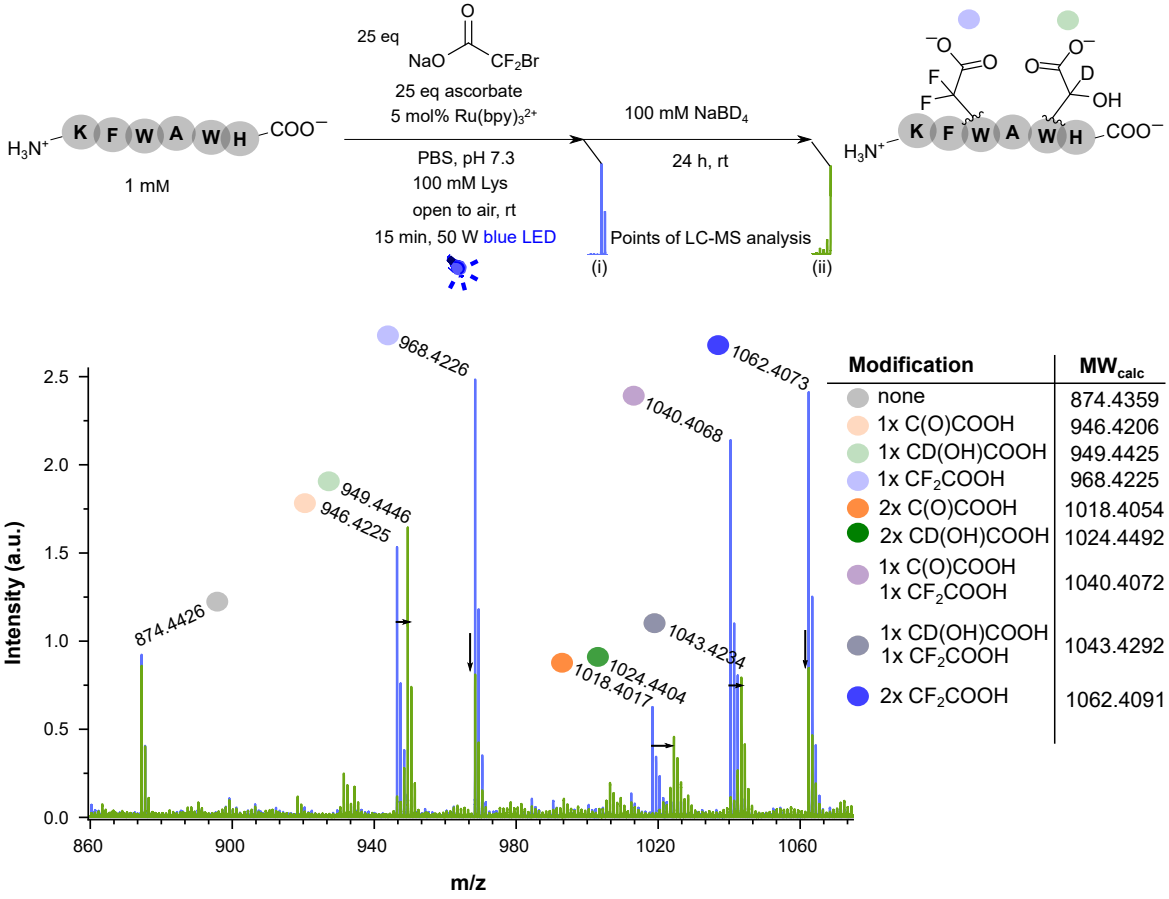


Fig. 6.6: Difluoroalkylation of the KFWAWH model peptide with $BrCF_2CO_2Na$ under photoredox activation and subsequent borodeuteride reduction assay. The combined mass spectra of the reaction mixture after irradiation (blue peaks), and following 24 h incubation upon the addition of $NaBD_4$ (green peaks), revealed expected reactivity patterns with the product peaks matching the given HR-MS calculated masses.

In order to verify the applicability of the perfluoalkyl bromides as precursors suitable for protein modifications, a more complex bromofluoroalkyl compound bearing a sugar moiety via amide linkage was tested in the reaction with myoglobin. Such type of compounds might be mimics of natural PTM of Trp residue - C2 mannosylation[388] - Fig. 6.7.

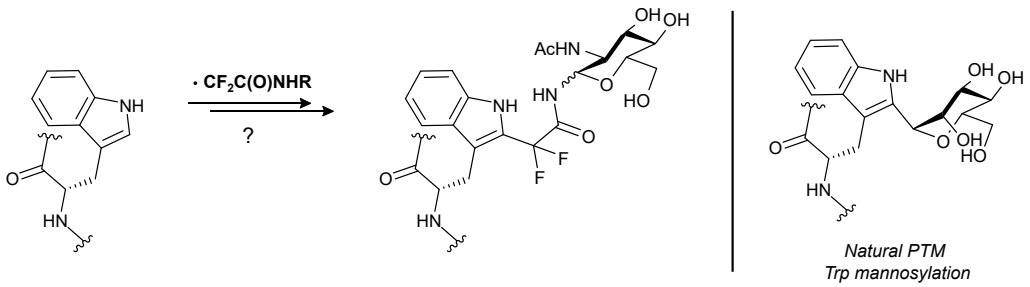


Fig. 6.7: Fluoroalkylation with a radical bearing sugar moiety and the structure of natural PTM - C2 mannosylated Trp.

Interestingly, this precursor allowed the formation of fluorinated products as well as small quantities of the acyl derivatives, as observed by LC-MS - Fig. 6.8. This result importantly demonstrated that the amide linkage is a suitable extension linker possibly allowing for the payload introduction, using a $CF_2C(O)NHR$ reactive radical.

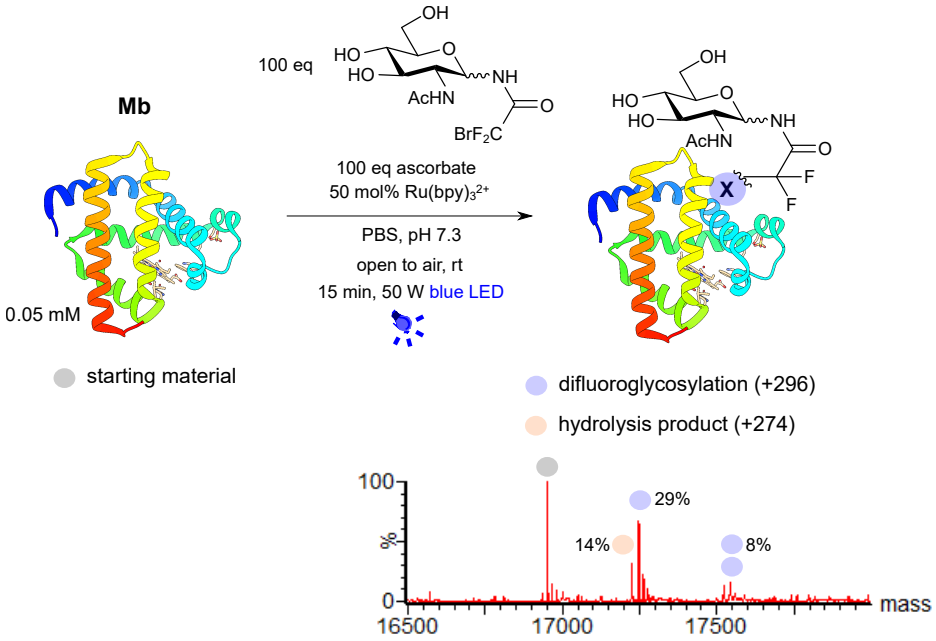


Fig. 6.8: Non-optimised difluoroglycosylation of myoglobin with the difluoroalkyl halide precursor under photoredox activation. Deconvoluted mass spectrum revealed the formation of the expected difluoroalkylation addition product and its hydrolysed derivative, in a single and double copy-number.

It is likely that the conditions could possibly be further optimised to achieve greater conversions, but as a general approach, the photoredox activation was validated as a potential platform for the further development of difluoroalkyl radical protein chemistry.

6.2 Conclusions

One of the main aims of this thesis was to assess the reactivity and selectivity of the chosen fluoroalkyl radicals in protein chemistry. This was described in detail for trifluoromethyl, difluoroethyl and difluoromethyl radicals in Chapters 2, 4 and 5, respectively. A brief perspective on a further panel of fluoroalkyl radicals was presented in Chapter 6 and recent literature reports have been discussed throughout the thesis, together forming a comprehensive view of the research topic. The main conclusion that could be drawn, was that the *fluoroalkyl* radicals are unique in their reactivity with native protein residues. Under comparable conditions *alkyl* radicals do not have a suitable acceptor across natural amino acid side chains (unpublished data Davis group). The presence of the fluorine atoms, promotes reactivity with a high selectivity for Trp in the cases of CF_3 and CF_2H radicals, and for other species and specific conditions, also with His and Tyr residues. An important feature of these radicals chemistry is the ability to modify residues within folded proteins. This was demonstrated by the effective modification of buried (hetero)aromatic side chains and their biophysical characterisation. This feature was utilised to chemically prepare proteins observable by ^{19}F NMR at aromatic residues and established a new approach in protein chemistry.

Although the main impact would be on the suggested chemical biology applications, one can imagine that the patterns of disclosed reactivity may also be conducive to research in material sciences and general bio-organic chemistry. Methods developed here may be applied for the introduction of fluorine for the purposes of physicochemical properties alteration, for example, in biomolecule stability engineering upon high level of fluorination.

Another open topic is radiofluorination of proteins; some approaches were suggested, and this work can be treated as a solid foundation for the further development of the idea.

A direct bioconjugation method using fluoroalkyl radicals was not established (one of the aims of this thesis), however, a even more powerful protocol, according to the 'Tag and Modify' strategy, with the reactive carbonyl tag derived from the difluoromethyl product was accomplished instead. In this way, this strategy tapped into a vast repertoire of aldehyde conjugation methods and simplify possible applications.

Finally, an under-explored aspect of this research is the photophysical investigation of the modified proteins. Targeting of Trp with minimal size and low perturbance tags resulted in altered emission characteristics. Further research (on determining the quantum yield, two-photon absorption cross section, fluorescence lifetime etc.) is needed to fully evaluate the potential of, now easily accessed, modified Trp residues in photobiology.

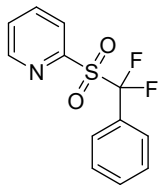
6.3 Experimental

The difluorobromo sugar radical precursor (*N*-(2-Acetamido-1,2-deoxy-D-glucosaminyl) bromodifluoroacetamide) was a kind gift of Dr Ben Bower, and the synthesis of which can be found elsewhere[399]. For further information regarding the experimental methods and general protocols (including photoredox modification of myoglobin) please refer to the relevant sections in Chapters 2, 4 and 5.

Myoglobin photoredox trifluoromethylation - feasibility studies

Standard photoredox fluoroalkylation conditions were adapted (Chapter 5), briefly: to a solution of a protein (97.5 μL, 513 μM in 100 mM of NH_4OAc), **BzSOOCF$_3$** ((0.1 - 1.5) μL, 0.1 M in DMSO), sodium ascorbate (0.3 μL, 1.65 M in the buffer) and $[Ru(bpy)_3]Cl_2 \cdot 6H_2O$ (0.76 μL, 3.3 mM in the buffer) were added. The reaction mixture was irradiated for 10 mins with blue LEDs (450 nm, 50 W) and after this, an aliquot (1 μL) was withdrawn and diluted with water (300 μL) for LC-MS analysis.

Synthesis of the PySO$_2$CF$_2$Ph radical precursor



Alkylation: PySH (2 g, 18 mmol) was dissolved in anhydrous MeCN (25 mL) under argon. Et_3N (3.8 mL, 27 mmol) was added to this solution. After 5 mins, benzyl bromide (2.65 mL, 21.7 mmol) was added dropwise over 5 mins. After starting material was consumed as indicated by TLC (~2 h), the reaction was quenched with HCl (20 mL, 1 M). The mixture was partitioned between EtOAc (60 mL) and water (15 mL), the aqueous phase was extracted with EtOAc (3 x 20 mL), the com-

bined organic phases were dried over $MgSO_4$ and concentrated in vacuo. Part of the residue was used in the next step without the purification.

Oxidation: Crude $PySCH_2Ph$ (0.5 g, ∼2.5 mmol) was dissolved in MeCN (15 mL) and DCM (12 mL). To this mixture aqueous KIO_4 (5.75 g, 20 mL, 25 mmol) suspension as well as $RuCl_3{\cdot}xH_2O$ (6 mg, ∼0.025 mmol) were added, and the reaction mixture was stirred overnight at rt. After this, the mixture was partitioned between DCM (40 mL) and water (10 mL), the layers separated and the aqueous phase was extracted with DCM (3 x 15 mL). Combined organic fractions were filtered, dried over $MgSO_4$, filtered through a silica plug and evaporated to dryness to yield a brownish solid that was used in the next step without the purification.

Fluorination: The crude $PySO_2CH_2Ph$ (227 mg, ∼1 mmol) was dissolved in anhydrous THF (39 mL) and cooled down to −70 °C. NaHMDS (2.2 mL of 1 M in THF, 2.2 mmol) was added dropwise over 15 mins. After this, NFSI (1.1 g, 3.5 mmol) was added in one portion. The reaction mixture was stirred for 120 mins at −70 °C and then warmed up to rt over 120 mins. The reaction was quenched with a saturated solution of NH_4Cl (20 mL) and partitioned between EtOAc (60 mL) and water (20 mL). The aqueous phase was extracted with EtOAc (3 x 25 mL), the organic phases were combined, washed with $NaHCO_3$ and brine, before drying over $MgSO_4$. The concentrated residue was purified using flash column chromatography using 4 g silica and the CombiFlash system with a gradient EtOAc in hexane (0% → 100%) to give **$PySO_2CF_2Ph$** as a white solid (81 mg, 0.3 mmol, ∼30%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.89 (dt, J = 4.6, 1.4 Hz, 1H, H6), 8.18 (d, J = 7.7 Hz, 1H, H3), 8.02 (td, J = 7.7, 1.4 Hz, 1H, H4), 7.72 (d, J = 8.0 Hz, 2H, H14&H10),

7.67 (ddd, J = 7.7, 4.6, 1.4 Hz, 1H, H5), 7.60 (d, J = 8.0 Hz, 1H, H12), 7.52 (t, J = 8.0 Hz, 2H, H11&H13),

^{13}**C NMR** (101 MHz, $CDCl_3$) δ 152.5 (C2), 151.0 (C6), 138.3 (C4), 132.7 (t, J = 1.8 Hz, C12), 128.7 (C13&C11), 128.7 (C5), 127.9 (t, J = 6.1 Hz, C10&C14), 126.7 (C3), 126.3 (t, J = 22.0 Hz, C9), 122.6 (t, J = 288.2 Hz, C8),

^{19}**F NMR** (470 MHz, $CDCl_3$) δ −100.7 (s, 2F),

HRMS-ESI(+): $[M+H]^+$: Calcd: 270.0395 found 270.0411.

Model peptide fluoroalkylations

To peptide solution (10 μL, 10 mM in PBS/Lys buffer), **$PySO_2CF_2Ph$** (2.5 μL, 25 eq, 1 M in DMSO), sodium ascorbate (1.5 μL, 25 eq, 1.65 M in PBS/Lys buffer), $[Ru(bpy)_3]^{2+}$ (1.5 μL, 5 mol%, 3.3 mM in PBS/Lys buffer) and PBS buffer (84.5 μL) were added. The reaction mixture was irradiated for 15 mins after which, an aliquot (2.5 μL) was diluted 10x with H_2O and analysed via LC-MS/MS. An identical procedure was applied when difluorobromoacetate was used as the radical precursor. After fluoroalkylation, $NaBD_4$ (0.4 mg) was added, and the reaction mixture was incubated for 24 h at rt when a final aliquot was analysed by LC-MS.

References

[1] B. Alberts, *Molecular Biology of the Cell - Chapter 3*, pp. 109–170. New York: Garland Science, 4th ed., 2002.

[2] F. Wold, "In Vivo Chemical Modification of Proteins (Post-Translational Modification)," *Annual Review of Biochemistry*, vol. 50, no. 1, pp. 783–814, 1981.

[3] D. G. Hoare and D. E. Koshland, "A Procedure for the Selective Modification of Carboxyl Groups in Proteins," *Journal of the American Chemical Society*, vol. 88, no. 9, pp. 2057–2058, 1966.

[4] L. Polgar and M. L. Bender, "A New Enzyme Containing a Synthetically Formed Active Site. Thiol-Subtilisin," *Journal of the American Chemical Society*, vol. 88, no. 13, pp. 3153–3154, 1966.

[5] K. E. Neet and D. E. Koshland, "The conversion of serine at the active site of subtilisin to cysteine: a 'chemical mutation'," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 56, no. 5, pp. 1606–1611, 1966.

[6] J. Armstrong and H. McKenzie, "A method for modification of carboxyl groups in proteins: Its application to the association of bovine β-lactoglobulin A," *Biochimica et Biophysica Acta - Protein Structure*, vol. 147, no. 1, pp. 93 – 99, 1967.

[7] K. N. Chuh, A. R. Batt, and M. R. Pratt, "Chemical Methods for Encoding and Decoding of Posttranslational Modifications," *Cell Chemical Biology*, vol. 23, no. 1, pp. 86–107, 2016.

[8] T. W. Muir, D. Sondhi, and P. A. Cole, "Expressed protein ligation: A general method for protein engineering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 12, pp. 6705–6710, 1998.

[9] P. Wang, S. W. Dong, J. H. Shieh, E. Peguero, R. Hendrickson, M. A. S. Moore, and S. J. Danishefsky, "Erythropoietin Derived by Chemical Synthesis," *Science*, vol. 342, no. 6164, pp. 1357–1360, 2013.

[10] M. D. Simon, F. X. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, and K. M. Shokat, "The site-specific installation of methyl-lysine analogs into recombinant histones," *Cell*, vol. 128, no. 5, pp. 1003–1012, 2007.

[11] S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, and B. G. Davis, "Expanding the diversity of chemical

protein modification allows post-translational mimicry," *Nature*, vol. 446, no. 7139, pp. 1105–1109, 2007.

[12] C. D. Spicer, T. Triemer, and B. G. Davis, "Palladium-Mediated Cell-Surface Labeling," *Journal of the American Chemical Society*, vol. 134, no. 2, pp. 800–803, 2012.

[13] N. J. Agard, J. A. Prescher, and C. R. Bertozzi, "A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems," *Journal of the American Chemical Society*, vol. 127, no. 31, pp. 11196–11196, 2005.

[14] Y. X. Chen, G. Triola, and H. Waldmann, "Bioorthogonal Chemistry for Site-Specific Labeling and Surface Immobilization of Proteins," *Accounts of Chemical Research*, vol. 44, no. 9, pp. 762–773, 2011.

[15] M. J. Birrer, K. N. Moore, I. Betella, and R. C. Bates, "Antibody-Drug Conjugate-Based Therapeutics: State of the Science," *Journal of the National Cancer Institute*, vol. 111, no. 6, pp. 538–549, 2019.

[16] D. G. Rawale, K. Thakur, S. R. Adusumalli, and V. Rai, "Chemical Methods for Selective Labeling of Proteins," *European Journal of Organic Chemistry*, vol. 2019, no. 40, pp. 6749–6763, 2019.

[17] S. Graslund, P. Nordlund, J. Weigelt, *et al.*, "Protein production and purification," *Nature Methods*, vol. 5, no. 2, pp. 135–146, 2008.

[18] J. M. Chalker, G. J. L. Bernardes, and B. G. Davis, "A 'Tag-and-Modify' Approach to Site-Selective Protein Modification," *Accounts of Chemical Research*, vol. 44, no. 9, pp. 730–741, 2011.

[19] P. E. Carrigan, P. Baller, and S. Tuzmen, "Site-Directed Mutagenesis," *Disease Gene Identification: Methods and Protocols*, vol. 700, pp. 107–124, 2011.

[20] J. C. M. van Hest and D. A. Tirrell, "Efficient introduction of alkene functionality into proteins in vivo," *FEBS Letters*, vol. 428, no. 1-2, pp. 68–70, 1998.

[21] K. Kirshenbaum, I. S. Carrico, and D. A. Tirrell, "Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues," *ChemBioChem*, vol. 3, no. 2-3, pp. 235–237, 2002.

[22] V. W. Cornish and P. G. Schultz, "A New Tool for Studying Protein-Structure and Function," *Current Opinion in Structural Biology*, vol. 4, no. 4, pp. 601–607, 1994.

[23] C. D. Spicer and B. G. Davis, "Selective chemical protein modification," *Nature Communications*, vol. 5, no. 4740, 2014.

[24] O. Boutureira and G. J. L. Bernardes, "Advances in Chemical Protein Modification," *Chemical Reviews*, vol. 115, no. 5, pp. 2174–2195, 2015.

[25] E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira, and G. J. L. Bernardes, "Contemporary approaches to site-selective protein modification," *Nature Reviews Chemistry*, vol. 3, no. 3, pp. 147–171, 2019.

[26] E. van Groesen, C. T. Lohans, J. Brem, K. M. J. Aertker, T. D. W. Claridge, and C. J. Schofield, "^{19}F NMR Monitoring of Reversible Protein Post-Translational Modifications: Class D β-Lactamase Carbamylation and Inhibition," *Chemistry - A European Journal*, vol. 25, no. 51, pp. 11837–11841, 2019.

[27] R. K. Lesniak, A. M. Rydzik, J. J. A. G. Kamps, A. Kahn, T. D. W. Claridge, and C. J. Schofield, "^{19}F NMR studies on γ-butyrobetaine hydroxylase provide mechanistic insights and suggest a dual inhibition mode," *Chemical Communications*, vol. 55, no. 98, pp. 14717–14720, 2019.

[28] A. Miseta and P. Csutora, "Relationship between the occurrence of cysteine in proteins and the complexity of organisms," *Molecular Biology and Evolution*, vol. 17, no. 8, pp. 1232–1239, 2000.

[29] M. Y. Yang, J. Li, and P. R. Chen, "Transition metal-mediated bioorthogonal protein chemistry in living cells," *Chemical Society Reviews*, vol. 43, no. 18, pp. 6511–6526, 2014.

[30] M. K. Bilyard, H. J. Bailey, L. Raich, M. A. Gafitescu, T. Machida, J. Iglesias-Fernandez, S. S. Lee, C. D. Spicer, C. Rovira, W. W. Yue, and B. G. Davis, "Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis," *Nature*, vol. 563, no. 7730, pp. 235–240, 2018.

[31] J. M. Chalker, C. S. C. Wood, and B. G. Davis, "A Convenient Catalyst for Aqueous and Protein Suzuki-Miyaura Cross-Coupling," *Journal of the American Chemical Society*, vol. 131, no. 45, pp. 16346–16347, 2009.

[32] R. J. Spears and M. A. Fascione, "Site-selective incorporation and ligation of protein aldehydes," *Organic & Biomolecular Chemistry*, vol. 14, no. 32, pp. 7622–7638, 2016.

[33] S. Pomplun, M. Y. H. Mohamed, T. Oelschlaegel, C. Wellner, and F. Bergmann, "Efficient Pictet-Spengler Bioconjugation with *N*-Substituted Pyrrolyl Alanine Derivatives," *Angewandte Chemie - International Edition*, vol. 58, no. 11, pp. 3542–3547, 2019.

[34] R. Walder, M. A. LeBlanc, W. J. Van Patten, D. T. Edwards, J. A. Greenberg, A. Adhikari, S. R. Okoniewski, R. M. A. Sullan, D. Rabuka, M. C. Sousa, and T. T. Perkins, "Rapid Characterization of a Mechanically Labile α-Helical Protein Enabled by Efficient Site-Specific Bioconjugation," *Journal of the American Chemical Society*, vol. 139, no. 29, pp. 9867–9875, 2017.

[35] X. H. Shi, Y. N. Jung, L. J. Lin, C. Liu, C. Wu, I. K. O. Cann, and T. Ha, "Quantitative Fluorescent Labeling of Aldehyde-Tagged Proteins for Single-Molecule Imaging," *Nature Methods*, vol. 9, no. 5, p. 499–503, 2012.

[36] A. Dumas, L. Lercher, C. D. Spicer, and B. G. Davis, "Designing logical codon reassignment - Expanding the chemistry in biology," *Chemical Science*, vol. 6, no. 1, pp. 50–69, 2015.

[37] T. P. Yang, X. M. Li, X. C. Bao, Y. M. E. Fung, and X. D. Li, "Photo-lysine captures proteins that bind lysine post-translational modifications," *Nature Chemical Biology*, vol. 12, no. 2, pp. 70–72, 2016.

[38] P. Thirumurugan, D. Matosiuk, and K. Jozwiak, "Click Chemistry for Drug Development and Diverse Chemical-Biology Applications," *Chemical Reviews*, vol. 113, no. 7, pp. 4905–4979, 2013.

[39] T. Ramenda, T. Kniess, R. Bergmann, J. Steinbach, and F. Wuest, "Radiolabelling of proteins with fluorine-18 via click chemistry," *Chemical Communications*, no. 48, pp. 7521–7523, 2009.

[40] Z. H. Li, D. L. Shen, S. Q. Hu, T. Su, K. Huang, F. R. Liu, L. Hou, and K. Cheng, "Pretargeting and Bioorthogonal Click Chemistry-Mediated Endogenous Stem Cell Homing for Heart Repair," *ACS Nano*, vol. 12, no. 12, pp. 12193–12200, 2018.

[41] S. T. Laughlin, J. M. Baskin, S. L. Amacher, and C. R. Bertozzi, "In vivo imaging of membrane-associated glycans in developing zebrafish," *Science*, vol. 320, no. 5876, pp. 664–667, 2008.

[42] H. Wang, R. B. Wang, K. M. Cai, H. He, Y. Liu, J. Yen, Z. Y. Wang, M. Xu, Y. W. Sun, X. Zhou, Q. Yin, L. Tang, I. T. Dobrucki, L. W. Dobrucki, E. J. Chaney, S. A. Boppart, T. M. Fan, S. Lezmi, X. S. Chen, L. C. Yin, and J. J. Cheng, "Selective *in vivo* metabolic cell-labeling-mediated cancer targeting," *Nature Chemical Biology*, vol. 13, no. 4, p. 415–424, 2017.

[43] G. J. L. Bernardes, L. Linderoth, K. J. Doores, O. Boutureira, and B. G. Davis, "Site-Selective Traceless Staudinger Ligation for Glycoprotein Synthesis Reveals Scope and Limitations," *ChemBioChem*, vol. 12, no. 9, pp. 1383–1386, 2011.

[44] E. Saxon, J. I. Armstrong, and C. R. Bertozzi, "A 'traceless' Staudinger ligation for the chemoselective synthesis of amide bonds," *Organic Letters*, vol. 2, no. 14, pp. 2141–2143, 2000.

[45] B. L. Oliveira, Z. Guo, and G. J. L. Bernardes, "Inverse electron demand Diels-Alder reactions in chemical biology," *Chemical Society Reviews*, vol. 46, no. 16, pp. 4895–4950, 2017.

[46] G. Zhang, J. Li, R. Xie, X. Y. Fan, Y. J. Liu, S. Q. Zheng, Y. Ge, and P. R. Chen, "Bioorthogonal Chemical Activation of Kinases in Living Systems," *ACS Central Science*, vol. 2, no. 5, pp. 325–331, 2016.

[47] Y. Y. A. Lin, O. Boutureira, L. Lercher, B. Bhushan, R. S. Paton, and B. G. Davis, "Rapid Cross-Metathesis for Reversible Protein Modifications via Chemical Access to Se-Allyl-selenocysteine in Proteins," *Journal of the American Chemical Society*, vol. 135, no. 33, pp. 12156–12159, 2013.

[48] E. M. Valkevich, R. G. Guenette, N. A. Sanchez, Y. C. Chen, Y. Ge, and E. R. Strieter, "Forging Isopeptide Bonds Using Thiol-Ene Chemistry: Site-Specific Coupling of Ubiquitin Molecules for Studying the Activity of Isopeptidases," *Journal of the American Chemical Society*, vol. 134, no. 16, pp. 6916–6919, 2012.

[49] J. Dadova, S. R. Galan, and B. G. Davis, "Synthesis of modified proteins via functionalization of dehydroalanine," *Current Opinion in Chemical Biology*, vol. 46, pp. 71–81, 2018.

[50] A. M. Freedy, M. J. Matos, O. Boutureira, F. Corzana, A. Guerreiro, P. Akkapeddi, V. J. Somovilla, T. Rodrigues, K. Nicholls, B. Xie, G. Jimenez-Oses, K. M. Brindle, A. A. Neves, and G. J. L. Bernardes, "Chemoselective Installation of Amine Bonds on Proteins through Aza-Michael Ligation," *Journal of the American Chemical Society*, vol. 139, no. 50, pp. 18365–18375, 2017.

[51] S. R. G. Galan, J. R. Wickens, J. Dadova, W. L. Ng, X. L. Zhang, R. A. Simion, R. Quinlan, E. Pires, R. S. Paton, S. Caddick, V. Chudasama, and B. G. Davis, "Post-translational site-selective protein backbone α-deuteration," *Nature Chemical Biology*, vol. 14, no. 10, p. 955–963, 2018.

[52] A. Yang, S. Ha, J. Ahn, R. Kim, S. Kim, Y. Lee, J. Kim, D. Soll, H. Y. Lee, and H. S. Park, "A chemical biology route to site-specific authentic protein modifications," *Science*, vol. 354, no. 6312, pp. 623–626, 2016.

[53] T. H. Wright, B. J. Bower, J. M. Chalker, G. J. L. Bernardes, R. Wiewiora, W. L. Ng, R. Raj, S. Faulkner, M. R. J. Vallee, A. Phanumartwiwath, O. D. Coleman, M. L. Thezenas, M. Khan, S. R. G. Galan, L. Lercher, M. W. Schombs, S. Gerstberger, M. E. Palm-Espling, A. J. Baldwin, B. M. Kessler, T. D. W. Claridge, S. Mohammed, and B. G. Davis, "Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity," *Science*, vol. 354, no. 6312, 2016.

[54] E. M. Pelegri-O'Day, E. W. Lin, and H. D. Maynard, "Therapeutic Protein-Polymer Conjugates: Advancing Beyond PEGylation," *Journal of the American Chemical Society*, vol. 136, no. 41, pp. 14323–14332, 2014.

[55] J. N. deGruyter, L. R. Malins, and P. S. Baran, "Residue-Specific Peptide Modification: A Chemist's Guide," *Biochemistry*, vol. 56, no. 30, pp. 3863–3873, 2017.

[56] W. Gong, G. Zhang, T. Liu, R. Giri, and J. Q. Yu, "Site-selective C(sp^3)-H functionalization of di-, tri-, and tetrapeptides at the N-terminus," *Journal of the American Chemical Society*, vol. 136, no. 48, pp. 16940–6, 2014.

[57] S. D. Maleknia and K. M. Downard, "Advances in radical probe mass spectrometry for protein footprinting in chemical biology applications," *Chemical Society Reviews*, vol. 43, no. 10, pp. 3244–3258, 2014.

[58] J. Ohata and Z. T. Ball, "A Hexa-rhodium Metallopeptide Catalyst for Site-Specific Functionalization of Natural Antibodies," *Journal of the American Chemical Society*, vol. 139, no. 36, pp. 12617–12622, 2017.

[59] S. Zheng and A. A. Doucette, "Preventing N- and O-formylation of proteins when incubated in concentrated formic acid," *Proteomics*, vol. 16, no. 7, pp. 1059–1068, 2016.

[60] B. V. Popp and Z. T. Ball, "Proximity-driven metallopeptide catalysis: Remarkable side-chain scope enables modification of the Fos bZip domain," *Chemical Science*, vol. 2, no. 4, pp. 690–695, 2011.

[61] J. C. Vantourout, S. R. Adusumalli, K. W. Knouse, D. T. Flood, A. Ramirez, N. M. Padial, A. Istrate, K. Maziarz, J. N. deGruyter, R. R. Merchant, J. X. Qiao, M. A. Schmidt, M. J. Deery, M. D. Eastgate, P. E. Dawson, G. J. L. Bernardes, and P. S. Baran, "Serine-Selective Bioconjugation," *Journal of the American Chemical Society*, no. Just Accepted, 2020.

[62] D. G. Hoare and D. E. Koshland, "A Method for Quantitative Modification and Estimation of Carboxylic Acid Groups in Proteins," *Journal of Biological Chemistry*, vol. 242, no. 10, pp. 2447–2453, 1967.

[63] N. A. McGrath, K. A. Andersen, A. K. F. Davis, J. E. Lomax, and R. T. Raines, "Diazo compounds for the bioreversible esterification of proteins," *Chemical Science*, vol. 6, no. 1, pp. 752–755, 2015.

[64] T. L. Schlick, Z. B. Ding, E. W. Kovacs, and M. B. Francis, "Dual-surface modification of the tobacco mosaic virus," *Journal of the American Chemical Society*, vol. 127, no. 11, pp. 3718–3723, 2005.

[65] O. Koniev and A. Wagner, "Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation," *Chemical Society Reviews*, vol. 44, no. 15, pp. 5743–5743, 2015.

[66] I. Dovgan, S. Erb, S. Hessmann, S. Ursuegui, C. Michel, C. Muller, G. Chaubet, S. Cianferani, and A. Wagner, "Arginine-selective bioconjugation with 4-azidophenyl glyoxal: application to the single and dual functionalisation of native antibodies," *Organic & Biomolecular Chemistry*, vol. 16, no. 8, pp. 1305–1311, 2018.

[67] J. Ohata, M. B. Minus, M. E. Abernathy, and Z. T. Ball, "Histidine-Directed Arylation/Alkenylation of Backbone N-H Bonds Mediated by Copper(II)," *Journal of the American Chemical Society*, vol. 138, no. 24, pp. 7472–7475, 2016.

[68] E. Hochuli, H. Dobeli, and A. Schacher, "New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighboring Histidine-Residues," *Journal of Chromatography*, vol. 411, pp. 177–184, 1987.

[69] S. Lata, M. Gavutis, R. Tampe, and J. Piehler, "Specific and stable fluorescence labeling of histidine-tagged proteins for dissecting multi-protein complex formation," *Journal of the American Chemical Society*, vol. 128, no. 7, pp. 2365–2372, 2006.

[70] G. Chen, A. Heim, D. Riether, D. Yee, Y. Milgrom, M. A. Gawinowicz, and D. Sames, "Reactivity of functional groups on the protein surface: Development of epoxide probes for protein labeling," *Journal of the American Chemical Society*, vol. 125, no. 27, pp. 8130–8133, 2003.

[71] Y. Takaoka, H. Tsutsumi, N. Kasagi, E. Nakata, and I. Hamachi, "One-pot and sequential organic chemistry on an enzyme surface to tether a fluorescent probe at the proximity of the active site with restoring enzyme activity," *Journal of the American Chemical Society*, vol. 128, no. 10, pp. 3273–3280, 2006.

[72] S. R. Adusumalli, T. G. Rawale, U. Singh, P. Tripathi, R. Paul, N. Kalra, R. K. Mishra, S. Shukla, and V. Rai, "Single-Site Labeling of Native Proteins Enabled by a Chemoselective and Site-Selective Chemical Technology," *Journal of the American Chemical Society*, vol. 140, no. 44, pp. 15114–15123, 2018.

[73] S. Jia, D. He, and C. J. Chang, "Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation," *Journal of the American Chemical Society*, vol. 141, no. 18, pp. 7294–7301, 2019.

[74] P. N. Joshi and V. Rai, "Single-site labeling of histidine in proteins, on-demand reversibility, and traceless metal-free protein purification," *Chemical Communications*, vol. 55, no. 8, pp. 1100–1103, 2019.

[75] X. P. Chen, F. R. Ye, X. S. Luo, X. Y. Liu, J. Zhao, S. Y. Wang, Q. Q. Zhou, G. Chen, and P. Wang, "Histidine-Specific Peptide Modification via Visible-Light-Promoted C-H Alkylation," *Journal of the American Chemical Society*, vol. 141, no. 45, pp. 18230–18237, 2019.

[76] K. Peciak, E. Laurine, R. Tommasi, J. W. Choi, and S. Brocchini, "Site-selective protein conjugation at histidine," *Chemical Science*, vol. 10, no. 2, pp. 427–439, 2019.

[77] A. I. Solomatina, P. S. Chelushkin, D. V. Krupenya, I. S. Podkorytov, T. O. Artamonova, V. V. Sizov, A. S. Melnikov, V. V. Gurzhiy, E. I. Koshel, V. I. Shcheslavskiy, and S. P. Tunik, "Coordination to Imidazole Ring Switches on Phosphorescence of Platinum Cyclometalated Complexes: The Route to Selective Labeling of Peptides and Proteins via Histidine Residues," *Bioconjugate Chemistry*, vol. 28, no. 2, pp. 426–437, 2017.

[78] A. F. M. Noisier, M. J. Johansson, L. Knerr, M. A. Hayes, W. J. Drury, E. Valeur, L. R. Malins, and R. Gopalakrishnan, "Late-Stage Functionalization of Histidine in Unprotected Peptides," *Angewandte Chemie - International Edition*, vol. 58, no. 52, pp. 19096–19102, 2019.

[79] L. H. Jones, A. Narayanan, and E. C. Hett, "Understanding and applying tyrosine biochemical diversity," *Molecular Biosystems*, vol. 10, no. 5, pp. 952–969, 2014.

[80] J. Riordan and B. Vallee, "*O*-Acetyltyrosine," *Methods in Enzymology*, vol. 25, no. C, pp. 500–506, 1972.

[81] Y. Koshi, E. Nakata, M. Miyagawa, S. Tsukiji, T. Ogawa, and I. Hamachi, "Target-specific chemical acylation of lectins by ligand-tethered DMAP catalysts," *Journal of the American Chemical Society*, vol. 130, no. 1, pp. 245–51, 2008.

[82] S. D. Tilley and M. B. Francis, "Tyrosine-selective protein alkylation using π-allylpalladium complexes," *Journal of the American Chemical Society*, vol. 128, no. 4, pp. 1080–1081, 2006.

[83] S. Sato and H. Nakamura, "Ligand-directed selective protein modification based on local single-electron-transfer catalysis," *Angewandte Chemie - International Edition*, vol. 52, no. 33, pp. 8681–4, 2013.

[84] H. A. Beard, J. R. Hauser, M. Walko, R. M. George, A. J. Wilson, and R. S. Bon, "Photocatalytic proximity labelling of MCL-1 by a BH3 ligand," *Communications Chemistry*, vol. 2, no. 133, 2019.

[85] K. L. Seim, A. C. Obermeyer, and M. B. Francis, "Oxidative Modification of Native Protein Residues Using Cerium(IV) Ammonium Nitrate," *Journal of the American Chemical Society*, vol. 133, no. 42, pp. 16970–16976, 2011.

[86] H. Ban, J. Gavrilyuk, and C. F. Barbas, "Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine," *Journal of the American Chemical Society*, vol. 132, no. 5, pp. 1523–1525, 2010.

[87] D. Alvarez-Dorta, C. Thobie-Gautier, M. Croyal, M. Bouzelha, M. Mevel, D. Deniaud, M. Boujtita, and S. G. Gouin, "Electrochemically Promoted Tyrosine-Click-Chemistry for Protein Labeling," *Journal of the American Chemical Society*, vol. 140, no. 49, pp. 17120–17126, 2018.

[88] J. Gavrilyuk, H. Ban, M. Nagano, W. Hakamata, and C. F. Barbas, "Formylbenzene Diazonium Hexafluorophosphate Reagent for Tyrosine-Selective Modification of Proteins and the Introduction of a Bioorthogonal Aldehyde," *Bioconjugate Chemistry*, vol. 23, no. 12, pp. 2321–2328, 2012.

[89] N. S. Joshi, L. R. Whitaker, and M. B. Francis, "A three-component Mannich-type reaction for selective tyrosine bioconjugation," *Journal of the American Chemical Society*, vol. 126, no. 49, pp. 15942–15943, 2004.

[90] J. M. McFarland, N. S. Joshi, and M. B. Francis, "Characterization of a three-component coupling reaction on proteins by isotopic labeling and nuclear magnetic resonance spectroscopy," *Journal of the American Chemical Society*, vol. 130, no. 24, pp. 7639–7644, 2008.

[91] J. Ohata, M. K. Miller, C. M. Mountain, F. Vohidov, and Z. T. Ball, "A Three-Component Organometallic Tyrosine Bioconjugation," *Angewandte Chemie - International Edition*, vol. 57, no. 11, pp. 2827–2830, 2018.

[92] C. L. Song, K. Liu, Z. J. Wang, B. Ding, S. C. Wang, Y. Weng, C. W. Chiang, and A. W. Lei, "Electrochemical oxidation induced selective tyrosine bioconjugation for the modification of biomolecules," *Chemical Science*, vol. 10, no. 34, pp. 7982–7987, 2019.

[93] A. W. Struck, M. R. Bennett, S. A. Shepherd, B. J. C. Law, Y. Zhuo, L. S. Wong, and J. Micklefield, "An Enzyme Cascade for Selective Modification of Tyrosine Residues in Structurally Diverse Peptides and Proteins," *Journal of the American Chemical Society*, vol. 138, no. 9, pp. 3038–3045, 2016.

[94] A. M. ElSohly and M. B. Francis, "Development of Oxidative Coupling Strategies for Site-Selective Protein Modification," *Accounts of Chemical Research*, vol. 48, no. 7, pp. 1971–1978, 2015.

[95] G. Espuna, D. Andreu, J. Barluenga, X. Perez, A. Planas, G. Arsequell, and G. Valencia, "Iodination of proteins by IPy_2BF_4, a new tool in protein chemistry," *Biochemistry*, vol. 45, no. 19, pp. 5957–5963, 2006.

[96] A. N. Glazer, "Specific Chemical Modification of Proteins," *Annual Review of Biochemistry*, vol. 39, pp. 101–130, 1970.

[97] J. R. Kramer and T. J. Deming, "Reversible chemoselective tagging and functionalization of methionine containing peptides," *Chemical Communications*, vol. 49, no. 45, pp. 5144–5146, 2013.

[98] M. T. Taylor, J. E. Nelson, M. G. Suero, and M. J. Gaunt, "A protein functionalization platform based on selective reactions at methionine residues," *Nature*, vol. 562, no. 7728, pp. 563–568, 2018.

[99] S. X. Lin, X. Y. Yang, S. Jia, A. M. Weeks, M. Hornsby, P. S. Lee, R. V. Nichiporuk, A. T. Iavarone, J. A. Wells, F. D. Toste, and C. J. Chang, "Redox-based reagents for chemoselective methionine bioconjugation," *Science*, vol. 355, no. 6325, pp. 597–602, 2017.

[100] J. Zang, Y. Chen, W. Zhu, and S. Lin, "Chemoselective Methionine Bioconjugation on a Polypeptide, Protein, and Proteome," *Biochemistry*, vol. 59, no. 2, pp. 132–138, 2020.

[101] S. Bloom, C. Liu, D. K. Kolmel, J. X. Qiao, Y. Zhang, M. A. Poss, W. R. Ewing, and D. W. C. MacMillan, "Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials," *Nature Chemistry*, vol. 10, no. 2, pp. 205–211, 2018.

[102] J. A. Shadish and C. A. DeForest, "Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials," *Matter*, vol. 2, no. 1, pp. 50–77, 2020.

[103] G. R. Grimsley, J. M. Scholtz, and C. N. Pace, "A summary of the measured pK_{a} values of the ionizable groups in folded proteins," *Protein Science*, vol. 18, no. 1, pp. 247–51, 2009.

[104] D. Chen, M. M. Disotuar, X. Xiong, Y. Wang, and D. H. Chou, "Selective N-terminal functionalization of native peptides and proteins," *Chemical Science*, vol. 8, no. 4, pp. 2717–2722, 2017.

[105] A. O. Y. Chan, C. M. Ho, H. C. Chong, Y. C. Leung, J. S. Huang, M. K. Wong, and C. M. Che, "Modification of N-Terminal α-Amino Groups of Peptides and Proteins Using Ketenes," *Journal of the American Chemical Society*, vol. 134, no. 5, pp. 2589–2598, 2012.

[106] J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi, and M. B. Francis, "N-terminal protein modification through a biomimetic transamination reaction," *Angewandte Chemie - International Edition*, vol. 45, no. 32, pp. 5307–5311, 2006.

[107] L. S. Witus, C. Netirojjanakul, K. S. Palla, E. M. Muehl, C.-H. Weng, A. T. Iavarone, and M. B. Francis, "Site-Specific Protein Transamination Using *N*-Methylpyridinium-4-carboxaldehyde," *Journal of the American Chemical Society*, vol. 135, no. 45, pp. 17223–17229, 2013.

[108] J. I. MacDonald, H. K. Munch, T. Moore, and M. B. Francis, "One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes," *Nature Chemical Biology*, vol. 11, no. 5, pp. 326–331, 2015.

[109] A. Onoda, N. Inoue, E. Sumiyoshi, and T. Hayashi, "Triazolecarbaldehyde Reagents for One-Step N-Terminal Protein Modification," *ChemBioChem*, vol. 21, no. 9, pp. 1274–1278, 2020.

[110] K. F. Geoghegan and J. G. Stroh, "Site-Directed Conjugation of Nonpeptide Groups to Peptides and Proteins Via Periodate-Oxidation of a 2-Amino Alcohol - Application to Modification at N-Terminal Serine," *Bioconjugate Chemistry*, vol. 3, no. 2, pp. 138–146, 1992.

[111] V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoet, J. C. M. Monbaliu, and O. Melnyk, "Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations," *Chemical Reviews*, vol. 119, no. 12, pp. 7328–7443, 2019.

[112] L. Purushottam, S. R. Adusumalli, U. Singh, V. B. Unnikrishnan, D. G. Rawale, M. Gujrati, R. K. Mishra, and V. Rai, "Single-site glycine-specific labeling of proteins," *Nature Communications*, vol. 10, no. 2539, 2019.

[113] C. Giglione, S. Fieulaine, and T. Meinnel, "N-terminal protein modifications: Bringing back into play the ribosome," *Biochimie*, vol. 114, pp. 134–146, 2015.

[114] S. D. Varfolomeev, I. V. Uporov, and E. V. Fedorov, "Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases," *Biochemistry (Moscow)*, vol. 67, no. 10, pp. 1099–1108, 2002.

[115] A. Fontana and C. Toniolo, "The chemistry of tryptophan in peptides and proteins," *Progress in the Chemistry of Organic Natural Products*, vol. 33, pp. 309–449, 1976.

[116] A. Foettinger, A. Leitner, and W. Lindner, "Selective enrichment of tryptophan-containing peptides from protein digests employing a reversible derivatization with malondialdehyde and solid-phase capture on hydrazide beads," *Journal of Proteome Research*, vol. 6, no. 9, pp. 3827–34, 2007.

[117] J. M. Antos and M. B. Francis, "Selective tryptophan modification with rhodium carbenoids in aqueous solution," *Journal of the American Chemical Society*, vol. 126, no. 33, pp. 10256–10257, 2004.

[118] J. M. Antos, J. M. McFarland, A. T. Iavarone, and M. B. Francis, "Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH," *Journal of the American Chemical Society*, vol. 131, no. 17, pp. 6301–6308, 2009.

[119] Z. Chen, B. V. Popp, C. L. Bovet, and Z. T. Ball, "Site-Specific Protein Modification with a Dirhodium Metallopeptide Catalyst," *ACS Chemical Biology*, vol. 6, no. 9, pp. 920–925, 2011.

[120] M. B. Hansen, F. Hubalek, T. Skrydstrup, and T. Hoeg-Jensen, "Chemo- and Regioselective Ethynylation of Tryptophan-Containing Peptides and Proteins," *Chemistry - A European Journal*, vol. 22, no. 5, pp. 1572–1576, 2016.

[121] Y. N. Yu, L. K. Zhang, A. V. Buevich, G. Q. Li, H. Q. Tang, P. Vachal, S. L. Colletti, and Z. C. Shi, "Chemoselective Peptide Modification via Photocatalytic Tryptophan β-Position Conjugation," *Journal of the American Chemical Society*, vol. 140, no. 22, pp. 6797–6800, 2018.

[122] Y. Seki, T. Ishiyama, D. Sasaki, J. Abe, Y. Sohma, K. Oisaki, and M. Kanai, "Transition metal-free tryptophan-selective bioconjugation of proteins," *Journal of the American Chemical Society*, vol. 138, no. 34, pp. 10798–10801, 2016.

[123] K. Rahimidashaghoul, I. Klimankova, M. Hubalek, M. Korecky, M. Chvojka, D. Pokorny, V. Matousek, L. Fojtik, D. Kavan, Z. Kukacka, P. Novak, and P. Beier, "Reductant-Induced Free Radical Fluoroalkylation of Nitrogen Heterocycles and Innate Aromatic Amino Acid Residues in Peptides and Proteins," *Chemistry - A European Journal*, vol. 25, no. 69, pp. 15779–15785, 2019.

[124] C. Perez-Rizquez, O. Abian, and J. M. Palomo, "Site-selective modification of tryptophan and protein tryptophan residues through PdNP bionanohybrid-catalysed C-H activation in aqueous media," *Chemical Communications*, vol. 55, no. 86, pp. 12928–12931, 2019.

[125] J. M. J. M. Ravasco, H. Faustino, A. Trindade, and P. M. P. Gois, "Bioconjugation with Maleimides: A Useful Tool for Chemical Biology," *Chemistry - A European Journal*, vol. 25, no. 1, pp. 43–59, 2019.

[126] Y. Q. Zhang, X. P. Zhou, Y. H. Xie, M. M. Greenberg, Z. Xi, and C. Z. Zhou, "Thiol Specific and Tracelessly Removable Bioconjugation via Michael Addition to 5-Methylene Pyrrolones," *Journal of the American Chemical Society*, vol. 139, no. 17, pp. 6146–6151, 2017.

[127] B. Bernardim, P. M. S. D. Cal, M. J. Matos, B. L. Oliveira, N. Martinez-Saez, I. S. Albuquerque, E. Perkins, F. Corzana, A. C. B. Burtoloso, G. Jimenez-Oses, and G. J. L. Bernardes, "Stoichiometric and irreversible cysteine-selective protein modification using carbonylacrylic reagents," *Nature Communications*, vol. 7, no. 13128, 2016.

[128] S. B. Gunnoo and A. Madder, "Chemical Protein Modification through Cysteine," *ChemBioChem*, vol. 17, no. 7, pp. 529–553, 2016.

[129] C. Zhang, M. Welborn, T. Y. Zhu, N. J. Yang, M. S. Santos, T. Van Voorhis, and B. L. Pentelute, "π-Clamp-mediated cysteine conjugation," *Nature Chemistry*, vol. 8, no. 2, pp. 120–128, 2016.

[130] C. P. Ramil, P. An, Z. Yu, and Q. Lin, "Sequence-Specific 2-Cyanobenzothiazole Ligation," *Journal of the American Chemical Society*, vol. 138, no. 17, pp. 5499–5502, 2016.

[131] C. Zhang, P. Dai, A. A. Vinogradov, Z. P. Gates, and B. L. Pentelute, "Site-Selective Cysteine-Cyclooctyne Conjugation," *Angewandte Chemie - International Edition*, vol. 57, no. 22, pp. 6459–6463, 2018.

[132] A. Dondoni, A. Massi, P. Nanni, and A. Roda, "A New Ligation Strategy for Peptide and Protein Glycosylation: Photoinduced Thiol-Ene Coupling," *Chemistry - A European Journal*, vol. 15, no. 43, pp. 11444–11449, 2009.

[133] E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, and S. L. Buchwald, "Organometallic palladium reagents for cysteine bioconjugation," *Nature*, vol. 526, no. 7575, pp. 687–691, 2015.

[134] V. H. Trang, E. M. Valkevich, S. Minami, Y. C. Chen, Y. Ge, and E. R. Strieter, "Nonenzymatic Polymerization of Ubiquitin: Single-Step Synthesis and Isolation of Discrete Ubiquitin Oligomers," *Angewandte Chemie - International Edition*, vol. 51, no. 52, pp. 13085–13088, 2012.

[135] C. Zhang, E. V. Vinogradova, A. M. Spokoyny, S. L. Buchwald, and B. L. Pentelute, "Arylation Chemistry for Bioconjugation," *Angewandte Chemie - International Edition*, vol. 58, no. 15, pp. 4810–4839, 2019.

[136] K. Kubota, P. Dai, B. L. Pentelute, and S. L. Buchwald, "Palladium Oxidative Addition Complexes for Peptide and Protein Cross-linking," *Journal of the American Chemical Society*, vol. 140, no. 8, pp. 3128–3133, 2018.

[137] M. S. Messina, J. M. Stauber, M. A. Waddington, A. L. Rheingold, H. D. Maynard, and A. M. Spokoyny, "Organometallic Gold(III) Reagents for Cysteine Arylation," *Journal of the American Chemical Society*, vol. 140, no. 23, pp. 7065–7069, 2018.

[138] T. Schlatzer, J. Kriegesmann, H. Schroder, M. Trobe, C. Lembacher-Fadum, S. Santner, A. V. Kravchuk, C. F. W. Becker, and R. Breinbauer, "Labeling and Natural Post-Translational Modification of Peptides and Proteins via Chemoselective Pd-Catalyzed Prenylation of Cysteine," *Journal of the American Chemical Society*, vol. 141, no. 37, pp. 14931–14937, 2019.

[139] L. Xu, M. Raabe, M. M. Zegota, J. Nogueira, V. Chudasama, S. L. Kuan, and T. Weil, "Site-selective protein modification via disulfide rebridging for fast tetrazine/trans-cyclooctene bioconjugation," *Organic & Biomolecular Chemistry*, vol. 18, pp. 1140–1147, 2020.

[140] G. Chaubet, F. Thoreau, and A. Wagner, "Recent, non-classical, approaches to antibody lysine modification," *Drug Discovery Today: Technologies*, vol. 30, pp. 21–26, 2018.

[141] K. Yamada and Y. Ito, "Recent Chemical Approaches for Site-Specific Conjugation of Native Antibodies: Technologies toward Next-Generation Antibody-Drug Conjugates," *ChemBioChem*, vol. 20, no. 21, pp. 2729–2737, 2019.

[142] D. G. Isom, C. A. Castaneda, B. R. Cannon, and B. E. Garcia-Moreno, "Large shifts in pK_a values of lysine residues buried inside a protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 13, pp. 5260–5265, 2011.

[143] M. J. Matos, B. L. Oliveira, N. Martinez-Saez, A. Guerreiro, P. M. S. D. Cal, J. Bertoldo, M. Maneiro, E. Perkins, J. Howard, M. J. Deery, J. M. Chalker, F. Corzana, G. Jimenez-Oses, and G. J. L. Bernardes, "Chemo- and Regioselective Lysine Modification on Native Proteins," *Journal of the American Chemical Society*, vol. 140, no. 11, pp. 4004–4017, 2018.

[144] B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy, and M. Howarth, "Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 12, pp. E690–E697, 2012.

[145] A. H. Keeble and M. Howarth, "Power to the protein: enhancing and combining activities using the Spy toolbox," *Chemical Science*, vol. 11, no. 28, pp. 7281–7291, 2020.

[146] S. Choi, S. Connelly, N. Reixach, I. A. Wilson, and J. W. Kelly, "Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma," *Nature Chemical Biology*, vol. 6, no. 2, pp. 133–139, 2010.

[147] Y. Zhang, Y. J. Liang, F. Huang, Y. Zhang, X. C. Li, and J. Xia, "Site-Selective Lysine Reactions Guided by Protein-Peptide Interaction," *Biochemistry*, vol. 58, no. 7, pp. 1010–1018, 2019.

[148] J. Mintseris and S. P. Gygi, "High-density chemical cross-linking for modeling protein interactions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 1, pp. 93–102, 2020.

[149] Q. Michaudel, Y. Ishihara, and P. S. Baran, "Academia-Industry Symbiosis in Organic Chemistry," *Accounts of Chemical Research*, vol. 48, no. 3, pp. 712–721, 2015.

[150] M. Yan, J. C. Lo, J. T. Edwards, and P. S. Baran, "Radicals: Reactive Intermediates with Translational Potential," *Journal of the American Chemical Society*, vol. 138, no. 39, pp. 12692–12714, 2016.

[151] S. Z. Zard, *Radical reactions in organic synthesis - Chapter 1 and 2*, pp. 1–47. Oxford Chemistry Masters, New York: Oxford University Press, 2003.

[152] J. Rittle and M. T. Green, "Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics," *Science*, vol. 330, no. 6006, pp. 933–937, 2010.

[153] B. J. Levin, Y. Y. Huang, S. C. Peck, Y. Wei, A. M. D. Campo, J. A. Marks, E. A. Franzosa, C. Huttenhower, and E. P. Balskus, "A prominent glycyl radical enzyme in human gut microbiomes metabolizes *trans*-4-hydroxy-L-proline," *Science*, vol. 355, no. 6325, 2017.

[154] B. C. Dickinson and C. J. Chang, "Chemistry and biology of reactive oxygen species in signaling or stress responses," *Nature Chemical Biology*, vol. 7, no. 8, pp. 504–11, 2011.

[155] F. De Vleeschouwer, V. Van Speybroeck, M. Waroquier, P. Geerlings, and F. De Proft, "Electrophilicity and nucleophilicity index for radicals," *Organic Letters*, vol. 9, no. 14, pp. 2721–2724, 2007.

[156] R. C. van Lier, A. D. de Bruijn, and G. Roelfes, "A Water-Soluble Iridium Photocatalyst for Chemical Modification of Dehydroalanines in Peptides and Proteins," *Chemistry – A European Journal*, no. Accepted Author Manuscript, 2020.

[157] N. Floyd, B. Vijayakrishnan, J. R. Koeppe, and B. G. Davis, "Thiyl glycosylation of olefinic proteins: S-linked glycoconjugate synthesis," *Angewandte Chemie - International Edition*, vol. 48, no. 42, pp. 7798–802, 2009.

[158] Y. Li, M. Pan, Y. Li, Y. Huang, and Q. Guo, "Thiol-yne radical reaction mediated site-specific protein labeling via genetic incorporation of an alkynyl-L-lysine analogue," *Organic & Biomolecular Chemistry*, vol. 11, no. 16, pp. 2624–9, 2013.

[159] M. Cheng, B. J. Zhang, W. D. Cui, and M. L. Gross, "Laser-Initiated Radical Trifluoromethylation of Peptides and Proteins: Application to Mass-Spectrometry-Based Protein Footprinting," *Angewandte Chemie - International Edition*, vol. 56, no. 45, pp. 14007–14010, 2017.

[160] E. R. Stadtman, "Oxidation of Free Amino-Acids and Amino-Acid-Residues in Proteins by Radiolysis and by Metal-Catalyzed Reactions," *Annual Review of Biochemistry*, vol. 62, pp. 797–821, 1993.

[161] J. D. Wallat, K. A. Rose, and J. K. Pokorski, "Proteins as substrates for controlled radical polymerization," *Polymer Chemistry*, vol. 5, no. 5, pp. 1545–1558, 2014.

[162] J. H. Ko and H. D. Maynard, "A guide to maximizing the therapeutic potential of protein-polymer conjugates by rational design," *Chemical Society Reviews*, vol. 47, no. 24, pp. 8998–9014, 2018.

[163] S. L. Baker, A. Munasinghe, B. Kaupbayeva, N. R. Kang, M. Certiat, H. Murata, K. Matyjaszewski, P. Lin, C. M. Colina, and A. J. Russell, "Transforming protein-polymer conjugate purification by tuning protein solubility," *Nature Communications*, vol. 10, no. 4718, 2019.

[164] C. J. Easton, "Free-radical reactions in the synthesis of α-amino acids and derivatives," *Chemical Reviews*, vol. 97, no. 1, pp. 53–82, 1997.

[165] M. J. Davies, "The oxidative environment and protein damage," *Biochimica Et Biophysica Acta - Proteins and Proteomics*, vol. 1703, no. 2, pp. 93–109, 2005.

[166] G. W. Preston and A. J. Wilson, "Photo-induced covalent cross-linking for the analysis of biomolecular interactions," *Chemical Society Reviews*, vol. 42, no. 8, pp. 3289–3301, 2013.

[167] G. Dorman and G. D. Prestwich, "Benzophenone photophores in biochemistry," *Biochemistry*, vol. 33, no. 19, pp. 5661–73, 1994.

[168] D. A. Fancy and T. Kodadek, "Chemistry for the analysis of protein-protein interactions: Rapid and efficient cross-linking triggered by long wavelength light," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 3, pp. 1317–1317, 2000.

[169] S. Sato, M. Tsushima, and H. Nakamura, "Target-protein-selective inactivation and labelling using an oxidative catalyst," *Organic & Biomolecular Chemistry*, vol. 16, no. 34, pp. 6168–6179, 2018.

[170] S. Sato and H. Nakamura, "Protein Chemical Labeling Using Biomimetic Radical Chemistry," *Molecules*, vol. 24, no. 3980, 2019.

[171] J. Stubbe and W. A. van der Donk, "Protein radicals in enzyme catalysis," *Chemical Reviews*, vol. 98, no. 2, pp. 705–762, 1998.

[172] C. L. Hawkins and M. J. Davies, "Generation and propagation of radical reactions on proteins," *Biochimica Et Biophysica Acta - Bioenergetics*, vol. 1504, no. 2-3, pp. 196–219, 2001.

[173] W. R. Dolbier, "Structure, reactivity, and chemistry of fluoroalkyl radicals," *Chemical Reviews*, vol. 96, no. 5, pp. 1557–1584, 1996.

[174] M. Yoshida, "Fluoroalkylated Styrene Dimers: Synthesis, Properties, and Applications," *Fluorinated Polymers, Vol 1: Synthesis, Properties, Processing and Simulation*, vol. 23, pp. 22–39, 2017.

[175] D. V. Avila, K. U. Ingold, J. Lusztyk, W. R. Dolbier, and H. Q. Pan, "Absolute rate constants for radical additions to alkenes in solution. The synergistic effect of perfluorination on the reactivities of *n*-alkyl radicals," *Journal of Organic Chemistry*, vol. 61, no. 6, pp. 2027–2030, 1996.

[176] H. X. Song, Q. Y. Han, C. L. Zhao, and C. P. Zhang, "Fluoroalkylation reactions in aqueous media: a review," *Green Chemistry*, vol. 20, no. 8, pp. 1662–1731, 2018.

[177] J. Rong, L. Deng, P. Tan, C. F. Ni, Y. C. Gu, and J. B. Hu, "Radical Fluoroalkylation of Isocyanides with Fluorinated Sulfones by Visible-Light Photoredox Catalysis," *Angewandte Chemie - International Edition*, vol. 55, no. 8, pp. 2743–2747, 2016.

[178] Y. N. Ji, T. Brueckl, R. D. Baxter, Y. Fujiwara, I. B. Seiple, S. Su, D. G. Blackmond, and P. S. Baran, "Innate C-H trifluoromethylation of heterocycles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 35, pp. 14411–14415, 2011.

[179] S. Barata-Vallejo, M. V. Cooke, and A. Postigo, "Radical Fluoroalkylation Reactions," *ACS Catalysis*, vol. 8, no. 8, pp. 7287–7307, 2018.

[180] D. A. Nagib and D. W. C. MacMillan, "Trifluoromethylation of arenes and heteroarenes by means of photoredox catalysis," *Nature*, vol. 480, no. 7376, pp. 224–228, 2011.

[181] D. E. Yerien, M. V. Cooke, M. C. G. Vior, S. Barata-Vallejo, and A. Postigo, "Radical fluoroalkylation reactions of (hetero) arenes and sulfides under red light photocatalysis," *Organic & Biomolecular Chemistry*, vol. 17, no. 15, pp. 3741–3746, 2019.

[182] W. Q. Hu, X. H. Xu, and F. L. Qing, "Visible light induced hydrodifluoromethylation of alkenes derived from oxindoles with (difluoromethyl)triphenylphosphonium bromide," *Journal of Fluorine Chemistry*, vol. 208, pp. 73–79, 2018.

[183] B. R. Langlois, E. Laurent, and N. Roidot, "Trifluoromethylation of Aromatic-Compounds with Sodium Trifluoromethanesulfinate under Oxidative Conditions," *Tetrahedron Letters*, vol. 32, no. 51, pp. 7525–7528, 1991.

[184] Y. Fujiwara, J. A. Dixon, F. O'Hara, E. D. Funder, D. D. Dixon, R. A. Rodriguez, R. D. Baxter, B. Herle, N. Sach, M. R. Collins, Y. Ishihara, and P. S. Baran, "Practical and innate carbon-hydrogen functionalization of heterocycles," *Nature*, vol. 492, no. 7427, pp. 95–99, 2012.

[185] S. M. Wang, J. B. Han, C. P. Zhang, H. L. Qin, and J. C. Xiao, "An overview of reductive trifluoromethylation reactions using electrophilic '$^{+}CF_3$' reagents," *Tetrahedron*, vol. 71, no. 42, pp. 7949–7976, 2015.

[186] X. Wang, Y. X. Ye, S. N. Zhang, J. J. Feng, Y. Xu, Y. Zhang, and J. B. Wang, "Copper-Catalyzed C(sp^3)-C(sp^3) Bond Formation Using a Hypervalent Iodine Reagent: An Efficient Allylic Trifluoromethylation," *Journal of the American Chemical Society*, vol. 133, no. 41, pp. 16410–16413, 2011.

[187] Y. Li and A. Studer, "Transition-Metal-Free Trifluoromethylaminoxylation of Alkenes," *Angewandte Chemie - International Edition*, vol. 51, no. 33, pp. 8221–8224, 2012.

[188] B. I. Usachev, "1-/2-/3-Fluoroalkyl-substituted indoles, promising medicinally and biologically beneficial compounds: Synthetic routes, significance and potential applications," *Journal of Fluorine Chemistry*, vol. 185, pp. 118–167, 2016.

[189] D. Ryzhakov, M. Jarret, R. Guillot, C. Kouklovsky, and G. Vincent, "Radical-Mediated Dearomatization of Indoles with Sulfinate Reagents for the Synthesis of Fluorinated Spirocyclic Indolines," *Organic Letters*, vol. 19, no. 23, pp. 6336–6339, 2017.

[190] M. A. J. Duncton, "Minisci reactions: Versatile CH-functionalizations for medicinal chemists," *RSC Medicinal Chemistry*, vol. 2, no. 12, pp. 1135–1161, 2011.

[191] C. Q. Lu, D. VanDerveer, and D. D. DesMarteau, "Fluoroalkylation of Imidazoles by Hypervalent Iodonium Salts," *Organic Letters*, vol. 10, no. 24, pp. 5565–5568, 2008.

[192] V. M. Labroo, R. B. Labroo, and L. A. Cohen, "Direct Photochemical Trifluoromethylation of Histidine-Containing Peptides," *Tetrahedron Letters*, vol. 31, no. 40, pp. 5705–5708, 1990.

[193] L. Li, X. Y. Mu, W. B. Liu, Y. C. Wang, Z. T. Mi, and C. J. Li, "Simple and Clean Photoinduced Aromatic Trifluoromethylation Reaction," *Journal of the American Chemical Society*, vol. 138, no. 18, pp. 5809–5812, 2016.

[194] N. Ichiishi, J. P. Caldwell, M. Lin, W. Zhong, X. H. Zhu, E. Streckfuss, H. Y. Kim, C. A. Parish, and S. W. Krska, "Protecting group free radical C-H trifluoromethylation of peptides," *Chemical Science*, vol. 9, no. 17, pp. 4168–4175, 2018.

[195] B. Langlois, D. Montegre, and N. Roidot, "Synthesis of S-Trifluoromethyl-Containing α-Amino-Acids from Sodium Trifluoromethanesulfinate and Dithio-Amino Acids," *Journal of Fluorine Chemistry*, vol. 68, no. 1, pp. 63–66, 1994.

[196] X. H. Xu, K. Matsuzaki, and N. Shibata, "Synthetic Methods for Compounds Having CF_3-S Units on Carbon by Trifluoromethylation, Trifluoromethylthiolation, Triflylation, and Related Reactions," *Chemical Reviews*, vol. 115, no. 2, pp. 731–764, 2015.

[197] J. J. Ma, Q. R. Liu, G. P. Lu, and W. B. Yi, "Silver-catalyzed fluoroalkylation of thiols using fluoroalkanesulfinates," *Journal of Fluorine Chemistry*, vol. 193, pp. 113–117, 2017.

[198] S. Capone, I. Kieltsch, O. Flogel, G. Lelais, A. Togni, and D. Seebach, "Electrophilic S-Trifluoromethylation of Cysteine Side Chains in α- and β-Peptides: Isolation of Trifluoro-methylated *Sandostatin* (Octreotide) Derivatives," *Helvetica Chimica Acta*, vol. 91, no. 11, pp. 2035–2056, 2008.

[199] K. K. J. Chan and D. O'Hagan, "The Rare Fluorinated Natural Products and Biotechnological Prospects for Fluorine Enzymology," *Natural Product Biosynthesis by Microorganisms and Plants, Part B*, vol. 516, pp. 219–235, 2012.

[200] S. Purser, P. R. Moore, S. Swallow, and V. Gouverneur, "Fluorine in medicinal chemistry," *Chemical Society Reviews*, vol. 37, no. 2, pp. 320–330, 2008.

[201] N. Leclerc, P. Chavez, O. A. Ibraikulov, T. Heiser, and P. Leveque, "Impact of Backbone Fluorination on π-Conjugated Polymers in Organic Photovoltaic Devices: A Review," *Polymers*, vol. 8, no. 11, 2016.

[202] M. Cheng, C. Y. Guo, and M. L. Gross, "The Application of Fluorine-Containing Reagents in Structural Proteomics," *Angewandte Chemie - International Edition*, vol. 59, no. 15, pp. 5880–5889, 2020.

[203] E. N. Marsh and Y. Suzuki, "Using ^{19}F NMR to probe biological interactions of proteins and peptides," *ACS Chemical Biology*, vol. 9, no. 6, pp. 1242–50, 2014.

[204] B. C. Buer and E. N. G. Marsh, "Fluorine: A new element in protein design," *Protein Science*, vol. 21, no. 4, pp. 453–462, 2012.

[205] S. J. Ye, B. Loll, A. A. Berger, U. Mulow, C. Alings, M. C. Wahl, and B. Koksch, "Fluorine teams up with water to restore inhibitor activity to mutant BPTI," *Chemical Science*, vol. 6, no. 9, pp. 5246–5254, 2015.

[206] S. Huhmann and B. Koksch, "Fine-Tuning the Proteolytic Stability of Peptides with Fluorinated Amino Acids," *European Journal of Organic Chemistry*, vol. 2018, no. 27-28, pp. 3667–3679, 2018.

[207] L. Merkel and N. Budisa, "Organic fluorine as a polypeptide building element: *in vivo* expression of fluorinated peptides, proteins and proteomes," *Organic & Biomolecular Chemistry*, vol. 10, no. 36, pp. 7241–7261, 2012.

[208] L. Wang, A. Brock, B. Herberich, and P. G. Schultz, "Expanding the genetic code of *Escherichia coli*," *Science*, vol. 292, no. 5516, pp. 498–500, 2001.

[209] J. C. Jackson, J. T. Hammill, and R. A. Mehl, "Site-specific incorporation of a ^{19}F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity," *Journal of the American Chemical Society*, vol. 129, no. 5, pp. 1160–1166, 2007.

[210] T. H. Kim, P. Mehrabi, Z. Ren, A. Sljoka, C. Ing, A. Bezginov, L. B. Ye, R. Pomes, R. S. Prosser, and E. F. Pai, "The role of dimer asymmetry and protomer dynamics in enzyme catalysis," *Science*, vol. 355, no. 6322, 2017.

[211] J. L. Kitevski-LeBlanc and R. S. Prosser, "Current applications of ^{19}F NMR to studies of protein structure and dynamics," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 62, pp. 1–33, 2012.

[212] L. B. Ye, N. Van Eps, M. Zimmer, O. P. Ernst, and R. S. Prosser, "Activation of the A(2A) adenosine G-protein-coupled receptor by conformational selection," *Nature*, vol. 533, no. 7602, pp. 265–268, 2016.

[213] J. Clark and D. O'Hagan, "Strategies for radiolabelling antibody, antibody fragments and affibodies with fluorine-18 as tracers for positron emission tomography (PET)," *Journal of Fluorine Chemistry*, vol. 203, pp. 31–46, 2017.

[214] O. Boutureira, G. J. L. Bernardes, F. D'Hooge, and B. G. Davis, "Direct radiolabelling of proteins at cysteine using [^{18}F]-fluorosugars," *Chemical Communications*, vol. 47, no. 36, pp. 10010–10012, 2011.

[215] H. S. Krishnan, L. L. Ma, N. Vasdev, and S. H. Liang, "^{18}F-Labeling of Sensitive Biomolecules for Positron Emission Tomography," *Chemistry - A European Journal*, vol. 23, no. 62, pp. 15553–15577, 2017.

[216] E. N. G. Marsh, "Fluorinated Proteins: From Design and Synthesis to Structure and Stability," *Accounts of Chemical Research*, vol. 47, no. 10, pp. 2878–2886, 2014.

[217] N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, and B. L. Pentelute, "Synthesis of proteins by automated flow chemistry," *Science*, vol. 368, no. 6494, pp. 980–987, 2020.

[218] E. N. G. Marsh, "Designing Fluorinated Proteins," *Peptide, Protein and Enzyme Design*, vol. 580, pp. 251–278, 2016.

[219] M. Salwiczek, E. K. Nyakatura, U. I. M. Gerling, S. J. Ye, and B. Koksch, "Fluorinated amino acids: compatibility with native protein structures and effects on protein-protein interactions," *Chemical Society Reviews*, vol. 41, no. 6, pp. 2135–2171, 2012.

[220] W. E. Hull and B. D. Sykes, "Fluorotyrosine alkaline phosphatase. ^{19}F nuclear magnetic resonance relaxation times and molecular motion of the individual fluorotyrosines," *Biochemistry*, vol. 13, no. 17, pp. 3431–7, 1974.

[221] H. Welte, T. Zhou, X. Mihajlenko, O. Mayans, and M. Kovermann, "What does fluorine do to a protein? Thermodynamic, and highly-resolved structural insights into fluorine-labelled variants of the cold shock protein," *Scientific Reports*, vol. 10, no. 1, p. 2640, 2020.

[222] C. Dalvit and A. Vulpetti, "Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects," *Journal of Medicinal Chemistry*, vol. 62, no. 5, pp. 2218–2244, 2019.

[223] C. T. Gee, K. E. Arntson, A. K. Urick, N. K. Mishra, L. M. L. Hawk, A. J. Wisniewski, and W. C. K. Pomerantz, "Protein-observed ^{19}F-NMR for fragment screening, affinity quantification and druggability assessment," *Nature Protocols*, vol. 11, no. 8, pp. 1414–1427, 2016.

[224] A. Boeszoermenyi, S. Chhabra, A. Dubey, D. L. Radeva, N. T. Burdzhiev, C. D. Chanev, O. I. Petrov, V. M. Gelev, M. Zhang, C. Anklin, H. Kovacs, G. Wagner, I. Kuprov, K. Takeuchi, and H. Arthanari, "Aromatic ^{19}F-^{13}C TROSY: a background-free approach to probe biomolecular structure, function, and dynamics," *Nature Methods*, vol. 16, no. 4, pp. 333–340, 2019.

[225] C. G. Li, G. F. Wang, Y. Q. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl, and G. J. Pielak, "Protein ^{19}F NMR in *Escherichia coli*," *Journal of the American Chemical Society*, vol. 132, no. 1, pp. 321–327, 2010.

[226] K. Kumar and A. Ghosh, "^{18}F-AIF Labeled Peptide and Protein Conjugates as Positron Emission Tomography Imaging Pharmaceuticals," *Bioconjugate Chemistry*, vol. 29, no. 4, pp. 953–975, 2018.

[227] C. Wangler, A. Kostikov, J. Zhu, J. Chin, B. Wangler, and R. Schirrmacher, "Silicon-[^{18}F] Fluorine Radiochemistry: Basics, Applications and Challenges," *Applied Sciences-Basel*, vol. 2, no. 2, pp. 277–302, 2012.

[228] K. E. Arntson and W. C. K. Pomerantz, "Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery," *Journal of Medicinal Chemistry*, vol. 59, no. 11, pp. 5158–5171, 2016.

[229] T. Tamura and I. Hamachi, "Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems," *Journal of the American Chemical Society*, vol. 141, no. 7, pp. 2782–2799, 2019.

[230] T. Didenko, J. J. Liu, R. Horst, R. C. Stevens, and K. Wuthrich, "Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs," *Current Opinion in Structural Biology*, vol. 23, no. 5, pp. 740–747, 2013.

[231] A. Studer, "A 'Renaissance' in Radical Trifluoromethylation," *Angewandte Chemie - International Edition*, vol. 51, no. 36, pp. 8950–8958, 2012.

[232] M. Imiolek, G. Karunanithy, W. L. Ng, A. J. Baldwin, V. Gouverneur, and B. G. Davis, "Selective Radical Trifluoromethylation of Native Residues in Proteins," *Journal of the American Chemical Society*, vol. 140, no. 5, pp. 1568–1571, 2018.

[233] B. Ding, Y. Weng, Y. Q. Liu, C. L. Song, L. Yin, J. F. Yuan, Y. R. Ren, A. W. Lei, and C. W. Chiang, "Selective Photoredox Trifluoromethylation of Tryptophan-Containing Peptides," *European Journal of Organic Chemistry*, vol. 2019, no. 46, pp. 7596–7605, 2019.

[234] I. Guerrero and A. Correa, "Cu-Catalyzed Site-Selective C(sp^2)-H Radical Trifluoromethylation of Tryptophan-Containing Peptides," *Organic Letters*, vol. 22, no. 5, pp. 1754–1759, 2020.

[235] J. M. Smith, J. A. Dixon, J. N. deGruyter, and P. S. Baran, "Alkyl Sulfinates: Radical Precursors Enabling Drug Discovery," *Journal of Medicinal Chemistry*, vol. 62, no. 5, pp. 2256–2264, 2019.

[236] W. R. Dolbier, *Guide to Fluorine NMR for Organic Chemists*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2nd ed., 2016.

[237] T. Bayer, A. Amberg, R. Bertermann, G. M. Rusch, M. W. Anders, and W. Dekant, "Biotransformation of 1,1,1,3,3-pentafluoropropane (HFC-245fa)," *Chemical Research in Toxicology*, vol. 15, no. 5, pp. 723–733, 2002.

[238] K. O. Christe and W. W. Wilson, "Nuclear Magnetic-Resonance Spectrum of the Fluoride Anion," *Journal of Fluorine Chemistry*, vol. 46, no. 2, pp. 339–342, 1990.

[239] A. L. Hansen and L. E. Kay, "Measurement of histidine pK_a values and tautomer populations in invisible protein states," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 17, pp. E1705–E1712, 2014.

[240] S. Rexroth, A. Poetsch, M. Rogner, A. Hamann, A. Werner, H. D. Osiewacz, E. R. Schafer, H. Seelert, and N. A. Dencher, "Reactive oxygen species target specific tryptophan site in the mitochondrial ATP synthase," *Biochimica Et Biophysica Acta - Bioenergetics*, vol. 1817, no. 2, pp. 381–387, 2012.

[241] X. M. Lam, J. Y. Yang, and J. L. Cleland, "Antioxidants for prevention of methionine oxidation in recombinant monoclonal antibody HER2," *Journal of Pharmaceutical Sciences*, vol. 86, no. 11, pp. 1250–1255, 1997.

[242] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, "Structure of Myoglobin - 3-Dimensional Fourier Synthesis at 2 Å Resolution," *Nature*, vol. 185, no. 4711, pp. 422–427, 1960.

[243] M. N. Chatterjea and R. Shinde, *Textbook of Medical Biochemistry - Section 2*, pp. 23–204. New Delhi: Jaypee Brothers Medical Publications (P) Ltd., 8th ed., 2012.

[244] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF chimera - A visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[245] C. J. Bendell, S. Liu, T. Aumentado-Armstrong, B. Istrate, P. T. Cernek, S. Khan, S. Picioreanu, M. Zhao, and R. A. Murgita, "Transient protein-protein interface prediction: datasets, features, algorithms, and the RAD-T predictor," *BMC Bioinformatics*, vol. 15, no. 82, 2014.

[246] S. Abbruzzetti, S. Sottini, C. Viappiani, and J. E. T. Corrie, "Acid-induced unfolding of myoglobin triggered by a laser pH jump method," *Photochemical & Photobiological Sciences*, vol. 5, no. 6, pp. 621–628, 2006.

[247] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations," *Nucleic Acids Research*, vol. 32, no. Web Server issue, pp. W665–7, 2004.

[248] J. M. Glandieres, C. Twist, A. Haouz, C. Zentz, and B. Alpert, "Resolved fluorescence of the two tryptophan residues in horse apomyoglobin," *Photochemistry and Photobiology*, vol. 71, no. 4, pp. 382–386, 2000.

[249] J. G. Pearson, B. Montez, H. B. Le, E. Oldfield, E. Y. T. Chien, and S. G. Sligar, "Assignment and analysis of fluorine nuclear magnetic resonance spectra of 4-fluorotryptophan myoglobins and hemoglobins," *Biochemistry*, vol. 36, no. 12, pp. 3590–3599, 1997.

[250] T. Yonetani and H. Schleyer, "Studies on Cytochrome C Peroxidase. IX. Reaction of Ferrimyoglobin with Hydroperoxides and a Comparison of Peroxide-Induced Compounds of Ferrimyoglobin and Cytochrome C Peroxidase," *Journal of Biological Chemistry*, vol. 242, no. 8, pp. 1974–1979, 1967.

[251] V. Katta and B. T. Chait, "Observation of the Heme Globin Complex in Native Myoglobin by Electrospray-Ionization Mass-Spectrometry," *Journal of the American Chemical Society*, vol. 113, no. 22, pp. 8534–8535, 1991.

[252] L. Wan, M. B. Twitchett, L. D. Eltis, A. G. Mauk, and M. Smith, "*In vitro* evolution of horse heart myoglobin to increase peroxidase activity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 22, pp. 12825–31, 1998.

[253] M. Fandrich, V. Forge, K. Buder, M. Kittler, C. M. Dobson, and S. Diekmann, "Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 26, pp. 15463–8, 2003.

[254] M. P. Richards, "Redox reactions of myoglobin," *Antioxidants & Redox Signaling*, vol. 18, no. 17, pp. 2342–51, 2013.

[255] A. Blumlein and J. J. McManus, "Reversible and non-reversible thermal denaturation of lysozyme with varying pH at low ionic strength," *Biochimica Et Biophysica Acta - Proteins and Proteomics*, vol. 1834, no. 10, pp. 2064–2070, 2013.

[256] P. Meers, "Location of Tryptophans in Membrane-Bound Annexins," *Biochemistry*, vol. 29, no. 13, pp. 3325–3330, 1990.

[257] N. O. Concha, J. F. Head, M. A. Kaetzel, J. R. Dedman, and B. A. Seaton, "Rat Annexin-V Crystal-Structure - Ca^{2+}-Induced Conformational-Changes," *Science*, vol. 261, no. 5126, pp. 1321–1324, 1993.

[258] W. C. Winkler, G. Gonzalez, J. B. Wittenberg, R. Hille, N. Dakappagari, A. Jacob, L. A. Gonzalez, and M. A. GillesGonzalez, "Nonsteric factors dominate binding of nitric oxide, azide, imidazole, cyanide, and fluoride to the Rhizobial heme-based oxygen sensor FixL," *Chemistry & Biology*, vol. 3, no. 10, pp. 841–850, 1996.

[259] S. Rahman, S. A. Ali, A. Islam, M. I. Hassan, and F. Ahmad, "Testing the dependence of stabilizing effect of osmolytes on the fractional increase in the accessible surface area on thermal and chemical denaturations of proteins," *Archives of Biochemistry and Biophysics*, vol. 591, pp. 7–17, 2016.

[260] X. Huang, C. X. Wang, L. R. Celeste, L. L. Lovelace, S. F. Sun, J. H. Dawson, and L. Lebioda, "Complex of myoglobin with phenol bound in a proximal cavity," *Acta Crystallographica Section F - Structural Biology Communications*, vol. 68, pp. 1465–1471, 2012.

[261] C. Jash and G. S. Kumar, "Binding of alkaloids berberine, palmatine and coralyne to lysozyme: a combined structural and thermodynamic study," *RSC Advances*, vol. 4, no. 24, pp. 12514–12525, 2014.

[262] J. P. Holland, "Chemical Kinetics of Radiolabelling Reactions," *Chemistry*, vol. 24, no. 62, pp. 16472–16483, 2018.

[263] I. Guerrero and A. Correa, "Site-Selective Trifluoromethylation Reactions of Oligopeptides," *Asian Journal of Organic Chemistry*, vol. 9, no. 6, pp. 898–909, 2020.

[264] S. Verhoog, C. W. Kee, Y. Wang, T. Khotavivattana, T. C. Wilson, V. Kersemans, S. Smart, M. Tredwell, B. G. Davis, and V. Gouverneur, "^{18}F-Trifluoromethylation of Unmodified Peptides with 5-^{18}F-(Trifluoromethyl)dibenzothiophenium Trifluoromethanesulfonate," *Journal of the American Chemical Society*, vol. 140, no. 5, pp. 1572–1575, 2018.

[265] M. Chrominski, M. R. Baranowski, S. Chmielinski, J. Kowalska, and J. Jemielity, "Synthesis of Trifluoromethylated Purine Ribonucleotides and Their Evaluation as ^{19}F NMR Probes," *The Journal of Organic Chemistry*, vol. 85, no. 5, pp. 3440–3453, 2020.

[266] A. Schischko, N. Kaplaneris, T. Rogge, G. Sirvinskaite, J. Son, and L. Ackermann, "Late-stage peptide C-H alkylation for bioorthogonal C-H activation featuring solid phase peptide synthesis," *Nature Communications*, vol. 10, no. 3553, 2019.

[267] F. Baud and S. Karlin, "Measures of residue density in protein structures," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 22, pp. 12494–12499, 1999.

[268] A. A. Berger, J. S. Voller, N. Budisa, and B. Koksch, "Deciphering the Fluorine Code - The Many Hats Fluorine Wears in a Protein Environment," *Accounts of Chemical Research*, vol. 50, no. 9, pp. 2093–2103, 2017.

[269] U. I. M. Gerling, M. Salwiczek, C. D. Cadicamo, H. Erdbrink, C. Czekelius, S. L. Grage, P. Wadhwani, A. S. Ulrich, M. Behrends, G. Haufe, and B. Koksch, "Fluorinated amino acids in amyloid formation: a symphony of size, hydrophobicity and α-helix propensity," *Chemical Science*, vol. 5, no. 2, pp. 819–830, 2014.

[270] S. Chowdhary, J. Moschner, D. J. Mikolajczak, M. Becker, A. F. Thunemann, C. Kastner, D. Klemczak, A. K. Stegemann, C. Bottcher, P. Metrangolo, R. R. Netz, and B. Koksch, "The Impact of Halogenated Phenylalanine Derivatives on NFGAIL Amyloid Formation," *ChemBioChem*, vol. 21, no. 24, pp. 3544–3554, 2020.

[271] W. M. Yau, W. C. Wimley, K. Gawrisch, and S. H. White, "The preference of tryptophan for membrane interfaces," *Biochemistry*, vol. 37, no. 42, pp. 14713–14718, 1998.

[272] L. P. Kozlowski, "Proteome-pI: proteome isoelectric point database," *Nucleic Acids Research*, vol. 45, no. D1, pp. D1112–D1116, 2017.

[273] J. Dadova, K. J. Wu, P. G. Isenegger, J. C. Errey, G. J. L. Bernardes, J. M. Chalker, L. Raich, C. Rovira, and B. G. Davis, "Precise Probing of Residue Roles by Post-Translational β,γ-C,N Aza-Michael Mutagenesis in Enzyme Active Sites," *ACS Central Science*, vol. 3, no. 11, pp. 1168–1173, 2017.

[274] G. De Sanctis, G. Petrella, C. Ciaccio, A. Feis, G. Smulevich, and M. Coletta, "A comparative study on axial coordination and ligand binding in ferric mini myoglobin and horse heart myoglobin," *Biophysical Journal*, vol. 93, no. 6, pp. 2135–42, 2007.

[275] M. R. Wilkins, E. Gasteiger, A. Bairoch, J.-C. Sanchez, K. L. Williams, R. D. Appel, and D. F. Hochstrasser, *Protein Identification and Analysis Tools in the ExPASy Server*, pp. 531–552. Totowa, NJ: Humana Press, 1999.

[276] J. Willwacher, R. Raj, S. Mohammed, and B. G. Davis, "Selective Metal-Site-Guided Arylation of Proteins," *Journal of the American Chemical Society*, vol. 138, no. 28, pp. 8678–81, 2016.

[277] H. Yasui, T. Yamamoto, E. Tokunaga, and N. Shibata, "Robust synthesis of trifluoromethionine and its derivatives by reductive trifluoromethylation of amino acid disulfides by CF_3I/Na/Liq.NH_3 system," *Journal of Fluorine Chemistry*, vol. 132, no. 3, pp. 186–189, 2011.

[278] K. B. Davis, Z. H. Zhang, E. A. Karpova, and J. Zhang, "Application of tyrosine-tryptophan fluorescence resonance energy transfer in monitoring protein size changes," *Analytical Biochemistry*, vol. 557, pp. 142–150, 2018.

[279] D. Schweida, P. Barraud, C. Regl, F. E. Loughlin, C. G. Huber, C. Cabrele, and M. Schubert, "The NMR signature of gluconoylation: a frequent N-terminal modification of isotope-labeled proteins," *Journal of Biomolecular NMR*, vol. 73, no. 1-2, pp. 71–79, 2019.

[280] D. Shugar, "The measurement of lysozyme activity and the ultra-violet inactivation of lysozyme," *Biochimica et Biophysica Acta*, vol. 8, no. 3, pp. 302–9, 1952.

[281] H. Guyon, H. Chachignon, and D. Cahard, "CF_3SO_2X (X = Na, Cl) as reagents for trifluoromethylation, trifluoromethylsulfenyl-, -sulfinyl- and -sulfonylation. Part 1: Use of CF_3SO_2Na," *Beilstein Journal of Organic Chemistry*, vol. 13, pp. 2764–2799, 2017.

[282] C. Zhang, "Application of Langlois' Reagent in Trifluoromethylation Reactions," *Advanced Synthesis & Catalysis*, vol. 356, no. 14-15, pp. 2895–2906, 2014.

[283] K. Muller, C. Faeh, and F. Diederich, "Fluorine in pharmaceuticals: Looking beyond intuition," *Science*, vol. 317, no. 5846, pp. 1881–1886, 2007.

[284] A. Bishop, N. Satyamurthy, G. Bida, and J. R. Barrio, "Chemical reactivity of the ^{18}F electrophilic reagents from the $^{18}O(p,n)^{18}F$ gas target systems," *Nuclear Medicine and Biology*, vol. 23, no. 5, pp. 559–565, 1996.

[285] S. Preshlock, M. Tredwell, and V. Gouverneur, "^{18}F-Labeling of Arenes and Heteroarenes for Applications in Positron Emission Tomography," *Chemical Reviews*, vol. 116, no. 2, pp. 719–766, 2016.

[286] V. T. Lien and P. J. Riss, "Radiosynthesis of [^{18}F]Trifluoroalkyl Groups: Scope and Limitations," *BioMed Research International*, no. 380124, p. 10, 2014.

[287] G. Angelini, M. Speranza, C. Y. Shiue, and A. P. Wolf, "$H^{18}F$ + Sb_2O_3 - a New Selective Radiofluorinating Agent," *Journal of the Chemical Society, Chemical Communications*, no. 12, pp. 924–925, 1986.

[288] T. Keller, F. R. Lopez-Picon, A. Krzyczmonik, S. Forsback, J. S. Takkinen, J. Rajander, S. Teperi, F. Dolle, J. O. Rinne, M. Haaparanta-Solin, and O. Solin, "Comparison of high and low molar activity TSPO tracer [^{18}F]F-DPA in a mouse model of Alzheimer's disease," *Journal of Cerebral Blood Flow & Metabolism*, vol. 40, no. 5, pp. 1012–1020, 2020.

[289] A. Wurzer, J. Pollmann, A. Schmidt, D. Reich, H. J. Wester, and J. Notni, "Molar Activity of Ga-68 Labeled PSMA Inhibitor Conjugates Determines PET Imaging Results," *Molecular Pharmaceutics*, vol. 15, no. 9, pp. 4296–4302, 2018.

[290] M. Suehiro, G. B. Yang, G. Torchon, E. Ackerstaff, J. Humm, J. Koutcher, and O. Ouerfelli, "Radiosynthesis of the tumor hypoxia marker [^{18}]TFMISO via *O*-[^{18}F]trifluoroethylation reveals a striking difference between trifluoroethyl tosylate and iodide in regiochemical reactivity toward oxygen nucleophiles," *Bioorganic & Medicinal Chemistry*, vol. 19, no. 7, pp. 2287–2297, 2011.

[291] P. S. Johnstrom and S. Stone-Elander, "Strategies for reducing isotopic dilution in the synthesis of ^{18}F-labeled polyfluorinated ethyl groups," *Applied Radiation and Isotopes*, vol. 47, no. 4, pp. 401–407, 1996.

[292] W. E. Truce and A. M. Murphy, "The Preparation of Sulfinic Acids," *Chemical Reviews*, vol. 48, no. 1, pp. 69–124, 1951.

[293] M. Tordeux, B. Langlois, and C. Wakselman, "Reactions of Bromotrifluoromethane and Related Halides. 8. Condensations with Dithionite and Hydroxymethanesulfinate Salts," *Journal of Organic Chemistry*, vol. 54, no. 10, pp. 2452–2453, 1989.

[294] X. Creary, "Nucleofugality of the Sulfinate Group in Carbocation-Forming Processes," *Journal of Organic Chemistry*, vol. 50, no. 25, pp. 5080–5084, 1985.

[295] Y. Ueno, A. Kojima, and M. Okawara, "A Convenient and General-Synthesis of Alkane Sulfinic Acids," *Chemistry Letters*, no. 12, pp. 2125–2128, 1984.

[296] E. V. Anslyn and D. A. Dougherty, *Modern Physical Organic Chemistry - Chapter 11*, pp. 627–705. Sausalito, CA: University Science, 2006.

[297] F. G. Bordwell and B. B. Jarvis, "Nucleophilic Displacements α to Sulfonyl Groupings," *Journal of Organic Chemistry*, vol. 33, no. 3, pp. 1182–1185, 1968.

[298] T. Khotavivattana, S. Verhoog, M. Tredwell, L. Pfeifer, S. Calderwood, K. Wheelhouse, T. L. Collier, and V. Gouverneur, "^{18}F-Labeling of Aryl-SCF_3, -OCF_3 and -$OCHF_2$ with [^{18}F]Fluoride," *Angewandte Chemie - International Edition*, vol. 54, no. 34, pp. 9991–9995, 2015.

[299] R. Gianatassio, S. Kawamura, C. L. Eprile, K. Foo, J. Ge, A. C. Burns, M. R. Collins, and P. S. Baran, "Simple sulfinate synthesis enables C-H trifluoromethylcyclopropanation," *Angewandte Chemie - International Edition*, vol. 53, no. 37, pp. 9851–5, 2014.

[300] B. Dhakal, L. Bohe, and D. Crich, "Trifluoromethanesulfonate Anion as Nucleophile in Organic Chemistry," *Journal of Organic Chemistry*, vol. 82, no. 18, pp. 9263–9269, 2017.

[301] P. Satori and C. Jünger, "Die elektrochemische Fluorierung von Alkansulfonamiden und Alkandisulfonamiden," *Journal of Fluorine Chemistry*, vol. 79, no. 1, pp. 71–75, 1996.

[302] N. Ignat'ev, A. Kucherina, and P. Sartori, "Comparative electrochemical fluorination of ethanesulfonyl chloride and fluoride," *Acta Chemica Scandinavica*, vol. 53, no. 12, pp. 1110–1116, 1999.

[303] A. A. Kolomeitsev, M. Vorobyev, and H. Gillandt, "Versatile application of trifluoromethyl triflate," *Tetrahedron Letters*, vol. 49, no. 3, pp. 449–454, 2008.

[304] G. A. Olah and T. Ohyama, "The Simple Practical Preparation of Trifluoromethyl Trifluoromethanesulfonate (Triflate)," *Synthesis*, vol. 1976, no. 05, pp. 319–320, 1976.

[305] T. Furuya, C. A. Kuttruff, and T. Ritter, "Carbon-fluorine bond formation," *Current Opinion In Drug Discovery & Development*, vol. 11, no. 6, p. 803—819, 2008.

[306] E. Carbonnel, T. Besset, T. Poisson, D. Labar, X. Pannecoucke, and P. Jubault, "^{18}F-Fluoroform: a ^{18}F-trifluoromethylating agent for the synthesis of $SCF_2^{18}F$-aromatic derivatives," *Chemical Communications*, vol. 53, no. 42, pp. 5706–5709, 2017.

[307] G. K. S. Prakash, F. Wang, Z. Zhang, R. Haiges, M. Rahm, K. O. Christe, T. Mathew, and G. A. Olah, "Long-Lived Trifluoromethanide Anion: A Key Intermediate in Nucleophilic Trifluoromethylations," *Angewandte Chemie - International Edition*, vol. 53, no. 43, pp. 11575–11578, 2014.

[308] M. Huiban, M. Tredwell, S. Mizuta, Z. H. Wan, X. M. Zhang, T. L. Collier, V. Gouverneur, and J. Passchier, "A broadly applicable [^{18}F]trifluoromethylation of aryl and heteroaryl iodides for PET imaging," *Nature Chemistry*, vol. 5, no. 11, pp. 941–944, 2013.

[309] S. S. Ashirbaev, V. V. Levin, M. I. Struchkova, and A. D. Dilman, "Coupling of *gem*-difluorinated organozinc reagents with S-electrophiles," *Journal of Fluorine Chemistry*, vol. 191, pp. 143–148, 2016.

[310] A. D. Dilman and V. V. Levin, "Difluorocarbene as a Building Block for Consecutive Bond-Forming Reactions," *Accounts of Chemical Research*, vol. 51, no. 5, pp. 1272–1280, 2018.

[311] J. G. Macneil and D. J. Burton, "Generation of Trifluoromethylcopper from Chlorodifluoroacetate," *Journal of Fluorine Chemistry*, vol. 55, no. 2, pp. 225–227, 1991.

[312] V. P. Mehta and M. F. Greaney, "S-, N-, and Se-Difluoromethylation Using Sodium Chlorodifluoroacetate," *Organic Letters*, vol. 15, no. 19, pp. 5036–5039, 2013.

[313] A. Lishchynskyi, M. A. Novikov, E. Martin, E. C. Escudero-Adan, P. Novak, and V. V. Grushin, "Trifluoromethylation of Aryl and Heteroaryl Halides with Fluoroform-Derived $CuCF_3$: Scope, Limitations, and Mechanistic Features," *Journal of Organic Chemistry*, vol. 78, no. 22, pp. 11126–11146, 2013.

[314] J. B. Geri, M. M. W. Wolfe, and N. K. Szymczak, "Borazine-CF_3- Adducts for Rapid, Room Temperature, and Broad Scope Trifluoromethylation," *Angewandte Chemie - International Edition*, vol. 57, no. 5, pp. 1381–1385, 2018.

[315] Y. C. Zhao, W. Z. Huang, L. G. Zhu, and J. B. Hu, "Difluoromethyl 2-Pyridyl Sulfone: A New *gem*-Difluoroolefination Reagent for Aldehydes and Ketones," *Organic Letters*, vol. 12, no. 7, pp. 1444–1447, 2010.

[316] S. Ou, C. R. Cao, M. Jiang, and J. T. Liu, "A Synthetic Approach to Tetrasubstituted Alkenes: Using β-Carbonyl Benzothiazol-2-yl Sulfones as Electrophiles," *European Journal of Organic Chemistry*, vol. 2013, no. 29, pp. 6510–6513, 2013.

[317] H. H. Coenen, B. Klatte, A. Knochel, M. Schuller, and G. Stocklin, "Preparation of N.C.A. [17-^{18}F]-Fluoroheptadecanoic Acid in High Yields Via Aminopolyether Supported, Nucleophilic Fluorination," *Journal of Labelled Compounds & Radiopharmaceuticals*, vol. 23, no. 5, pp. 455–466, 1986.

[318] J. Van den Brande, T. de Groot, K. Serdons, M. Bex, D. Ory, G. Bormans, and A. Verbruggen, "Retention of [^{18}F]fluoride on different reversed phase (RP) HPLC columns," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 41, pp. S243–S244, 2014.

[319] C. S. Thomoson and W. R. Dolbier, "Use of Fluoroform as a Source of Difluorocarbene in the Synthesis of Difluoromethoxy- and Difluorothiomethoxyarenes," *Journal of Organic Chemistry*, vol. 78, no. 17, pp. 8904–8908, 2013.

[320] C. W. Kee, O. Tack, F. Guibbal, T. C. Wilson, P. G. Isenegger, M. Imiolek, S. Verhoog, M. Tilby, G. Boscutti, S. Ashworth, J. Chupin, R. Kashani, A. W. J. Poh, J. K. Sosabowski, S. Macholl, C. Plisson, B. Cornelissen, M. C. Willis, J. Passchier, B. G. Davis, and V. Gouverneur, "^{18}F-Trifluoromethanesulfinate Enables Direct C-H ^{18}F-Trifluoromethylation of Native Aromatic Residues in Peptides," *Journal of the American Chemical Society*, vol. 142, no. 3, pp. 1180–1185, 2020.

[321] L. Trump, A. Lemos, B. Lallemand, P. Pasau, J. Mercier, C. Lemaire, A. Luxen, and C. Genicot, "Late-Stage ^{18}F-Difluoromethyl Labeling of N-Heteroaromatics with High Molar Activity for PET Imaging," *Angewandte Chemie - International Edition*, vol. 58, no. 37, pp. 13149–13154, 2019.

[322] B. Yang, S. Telu, M. B. Haskali, C. L. Morse, and V. W. Pike, "A Gas Phase Route to [^{18}F]fluoroform with Limited Molar Activity Dilution," *Scientific Reports*, vol. 9, no. 14835, 2019.

[323] V. A. Petrov, "A New Route to Polyfluorinated Trifluoromethanesulfonates - Synthesis of Perfluoroallyl and Perfluorobenzyl Triflates," *Journal of Fluorine Chemistry*, vol. 73, no. 1, pp. 17–19, 1995.

[324] M. Griesser, J. P. R. Chauvin, and D. A. Pratt, "The hydrogen atom transfer reactivity of sulfinic acids," *Chemical Science*, vol. 9, no. 36, pp. 7218–7229, 2018.

[325] C. G. Swain and C. B. Scott, "Rates of Solvolysis of Some Alkyl Fluorides and Chlorides," *Journal of the American Chemical Society*, vol. 75, no. 1, pp. 246–248, 1953.

[326] D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. G. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. P. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H. J. Vogel, and I. Forsythe, "HMDB: a knowledgebase for the human metabolome," *Nucleic Acids Research*, vol. 37, pp. D603–D610, 2009.

[327] G. Platzer, M. Okon, and L. P. McIntosh, "pH-dependent random coil ^{1}H, ^{13}C, and ^{15}N chemical shifts of the ionizable amino acids: a guide for protein pK_{a} measurements," *Journal of Biomolecular NMR*, vol. 60, no. 2-3, pp. 109–129, 2014.

[328] A. C. Conibear, K. J. Rosengren, C. F. W. Becker, and H. Kaehlig, "Random coil shifts of posttranslationally modified amino acids," *Journal of Biomolecular NMR*, vol. 73, no. 10-11, pp. 587–599, 2019.

[329] G. Prabhu, N. Narendra, Basavaprabhu, V. Panduranga, and V. V. Sureshbabu, "Amino acid fluorides: viable tools for synthesis of peptides, peptidomimetics and enantiopure heterocycles," *RSC Advances*, vol. 5, no. 60, pp. 48331–48362, 2015.

[330] H. O. Pritchard and H. A. Skinner, "The Heat of Hydrolysis of Acetyl Fluoride," *Journal of the Chemical Society*, no. Apr, pp. 1099–1099, 1950.

[331] M. O'Neil, ed., *The Merck Index - An Encyclopedia of Chemicals, Drugs, and Biologicals*, p. 977. New Jersey: Merck and Co., 2006.

[332] R. L. Schiltz, C. A. Mizzen, A. Vassilev, R. G. Cook, C. D. Allis, and Y. Nakatani, "Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates," *Journal of Biological Chemistry*, vol. 274, no. 3, pp. 1189–1192, 1999.

[333] E. M. Woolridge and S. E. Rokita, "Synthesis and Reactivity of 6-(Fluoromethyl)Indole and 6-(Difluoromethyl)Indole," *Tetrahedron Letters*, vol. 30, no. 45, pp. 6117–6120, 1989.

[334] E. Tuan and K. L. Kirk, "Fluorine reactivity in difluoromethylimidazoles," *Journal of Fluorine Chemistry*, vol. 127, no. 7, pp. 980–982, 2006.

[335] Q. H. Zhou, J. H. Gui, C. M. Pan, E. Albone, X. Cheng, E. M. Suh, L. Grasso, Y. Ishihara, and P. S. Baran, "Bioconjugation by Native Chemical Tagging of C-H Bonds," *Journal of the American Chemical Society*, vol. 135, no. 35, pp. 12994–12997, 2013.

[336] L. Horner and H. Neumann, "Studien Zum Vorgang Der Wasserstoffubertragung. XII. Hydrierende Spaltung Von Sulfonen Mit Tetramethylammonium Als Elektronenubertrager," *European Journal of Inorganic Chemistry*, vol. 98, no. 6, p. 1715, 1965.

[337] R. R. Merchant, J. T. Edwards, T. Qin, M. M. Kruszyk, C. Bi, G. D. Che, D. H. Bao, W. H. Qiao, L. J. Sun, M. R. Collins, O. O. Fadeyi, G. M. Gallego, J. J. Mousseau, P. Nuhant, and P. S. Baran, "Modular radical cross-coupling with sulfones enables access to sp^3-rich (fluoro)alkylated scaffolds," *Science*, vol. 360, no. 6384, pp. 75–80, 2018.

[338] S. Angerani and N. Winssinger, "Visible Light Photoredox Catalysis Using Ruthenium Complexes in Chemical Biology," *Chemistry - A European Journal*, vol. 25, no. 27, pp. 6661–6672, 2019.

[339] Y. Y. Chen, A. S. Kamlet, J. B. Steinman, and D. R. Liu, "A biomolecule-compatible visible-light-induced azide reduction from a DNA-encoded reaction-discovery system," *Nature Chemistry*, vol. 3, no. 2, pp. 146–153, 2011.

[340] C. M. Yang and C. U. Pittman, "Reductions of organic functional groups using $NaBH_4$ or $NaBH_4$/LiCl in diglyme at 125 to 162 °C," *Synthetic Communications*, vol. 28, no. 11, pp. 2027–2041, 1998.

[341] E. M. Sletten and C. R. Bertozzi, "Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality," *Angewandte Chemie - International Edition*, vol. 48, no. 38, pp. 6974–6998, 2009.

[342] J. L. Kice, G. Guaraldi, and C. G. Venier, "The Mechanism of the Disproportionation of Sulfinic Acids. Rate and Equilibrium Constants for the Sulfinic Acid-Sulfinyl Sulfone (Sulfinic Anhydride) Equilibrium," *The Journal of Organic Chemistry*, vol. 31, no. 11, pp. 3561–3567, 1966.

[343] I. Klimankova, M. Hubalek, V. Matousek, and P. Beier, "Synthesis of water-soluble hypervalent iodine reagents for fluoroalkylation of biological thiols," *Organic & Biomolecular Chemistry*, vol. 17, no. 47, pp. 10097–10102, 2019.

[344] T. Narita, B. T. Weinert, and C. Choudhary, "Functions and mechanisms of non-histone protein acetylation," *Nature Reviews Molecular Cell Biology*, vol. 20, no. 3, pp. 156–174, 2019.

[345] J. F. Riordan and B. L. Vallee, "[41] acetylation," *Methods in Enzymology*, vol. 25, pp. 494–9, 1972.

[346] J. Baeza, J. A. Dowell, M. J. Smallegan, J. Fan, D. Amador-Noguez, Z. Khan, and J. M. Denu, "Stoichiometry of Site-specific Lysine Acetylation in an Entire Proteome," *Journal of Biological Chemistry*, vol. 289, no. 31, pp. 21326–21338, 2014.

[347] S. Sidoli, Z. F. Yuan, S. Lin, K. Karch, X. S. Wang, N. Bhanu, A. M. Arnaudo, L. M. Britton, X. J. Cao, M. Gonzales-Cope, Y. M. Han, S. C. Liu, R. C. Molden, S. Wein, L. Afjehi-Sadat, and B. A. Garcia, "Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis," *Proteomics*, vol. 15, no. 9, pp. 1459–1469, 2015.

[348] K. S. Fritz, "Chemical Acetylation and Deacetylation," *Sirtuins: Methods and Protocols*, vol. 1077, pp. 191–201, 2013.

[349] C. Bunton and J. Fendler, "The hydrolysis of acetyl fluoride," *The Journal of Organic Chemistry*, vol. 31, no. 7, pp. 2307–2312, 1966.

[350] J. A. Devore and H. E. O'Neal, "Heats of formation of the acetyl halides and of the acetyl radical," *The Journal of Physical Chemistry*, vol. 73, no. 8, pp. 2644–2648, 1969.

[351] D. P. N. Satchell, "The Mechanism of Acylation by Carboxylic Anhydrides in Non-Aqueous Media, Catalysed by Mineral Acid," *Journal of the Chemical Society*, no. Apr, pp. 1752–1768, 1960.

[352] C. Fernandez-Garcia and M. W. Powner, "Selective Acylation of Nucleosides, Nucleotides, and Glycerol-3-phosphocholine in Water," *Synlett*, vol. 28, no. 1, pp. 78–83, 2017.

[353] A. R. Katritzky, K. Suzuki, and S. K. Singh, "N-acylation in combinatorial chemistry," *Arkivoc*, pp. 12–35, 2004.

[354] I. Dovgan, S. Ursuegui, S. Erb, C. Miche, S. Kolodych, S. Cianferani, and A. Wagner, "Acyl Fluorides: Fast, Efficient, and Versatile Lysine-Based Protein Conjugation via Plug-and-Play Strategy," *Bioconjugate Chemistry*, vol. 28, no. 5, pp. 1452–1457, 2017.

[355] F. O'Hara, D. G. Blackmond, and P. S. Baran, "Radical-based regioselective C-H functionalization of electron-deficient heteroarenes: scope, tunability, and predictability," *Journal of the American Chemical Society*, vol. 135, no. 32, pp. 12122–34, 2013.

[356] K. Rahimidashaghoul, I. Klimánková, M. Hubálek, V. Matoušek, J. Filgas, P. Slavíček, T. Slanina, and P. Beier, "Visible-light-driven fluoroalkylation of tryptophan residues in peptides," *ChemPhotoChem*, no. Accepted Author Manuscript, 2020.

[357] W. P. Jencks, "Mechanism and Catalysis of Simple Carbonyl Group Reactions," *Progress in Physical Organic Chemistry*, vol. 2, pp. 63–128, 1964.

[358] P. N. Dyer, R. S. Edayathumangalam, C. L. White, Y. H. Bao, S. Chakravarthy, U. M. Muthurajan, and K. Luger, "Reconstitution of nucleosome core particles from recombinant histones and DNA," *Chromatin and Chromatin Remodeling Enzymes, Part A*, vol. 375, pp. 23–44, 2004.

[359] Q. Zhou, A. Ruffoni, R. Gianatassio, Y. Fujiwara, E. Sella, D. Shabat, and P. S. Baran, "Direct synthesis of fluorinated heteroarylether bioisosteres," *Angewandte Chemie - International Edition*, vol. 52, no. 14, pp. 3949–3952, 2013.

[360] L. Wang, Z. W. Zhang, A. Brock, and P. G. Schultz, "Addition of the keto functional group to the genetic code of *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 1, pp. 56–61, 2003.

[361] M. R. Fleissner, E. M. Brustad, T. Kalai, C. Altenbach, D. Cascio, F. B. Peters, K. Hideg, S. Peuker, P. G. Schultz, and W. L. Hubbell, "Site-directed spin labeling of a genetically encoded unnatural amino acid," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 51, pp. 21637–21642, 2009.

[362] A. Tuley, Y. J. Lee, B. Wu, Z. U. Wang, and W. R. Liu, "A genetically encoded aldehyde for rapid protein labelling," *Chemical Communications*, vol. 50, no. 56, pp. 7424–7426, 2014.

[363] H. B. F. Dixon, "N-Terminal Modification of Proteins - a Review," *Journal of Protein Chemistry*, vol. 3, no. 1, pp. 99–108, 1984.

[364] M. Wilchek and E. A. Bayer, "Labeling Glycoconjugates with Hydrazide Reagents," *Methods in Enzymology*, vol. 138, pp. 429–442, 1987.

[365] I. S. Carrico, B. L. Carlson, and C. R. Bertozzi, "Introducing genetically encoded aldehydes into proteins," *Nature Chemical Biology*, vol. 3, no. 6, pp. 321–322, 2007.

[366] T. Kruger, T. Dierks, and N. Sewald, "Formylglycine-generating enzymes for site-specific bioconjugation," *Biological Chemistry*, vol. 400, no. 3, pp. 289–297, 2019.

[367] S. K. Jensen, S. R. Keiding, and J. Thogersen, "The hunt for HCO(aq)," *Physical Chemistry Chemical Physics*, vol. 12, no. 31, pp. 8926–8933, 2010.

[368] C. L. Ladner, K. Tran, M. Le, R. J. Turner, and R. A. Edwards, "Excited State Photoreaction between the Indole Side Chain of Tryptophan and Halocompounds Generates New Fluorophores and Unique Modifications," *Photochemistry and Photobiology*, vol. 90, no. 5, pp. 1027–1033, 2014.

[369] G. A. Olah, G. K. S. Prakash, Q. Wang, and X.-Y. Li, *Encyclopedia of Reagents for Organic Synthesis - Formyl Fluoride*, p. Online access. American Cancer Society, 2001.

[370] C. C. Blake, D. F. Koenig, G. A. Mair, A. C. North, D. C. Phillips, and V. R. Sarma, "Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution," *Nature*, vol. 206, no. 4986, pp. 757–61, 1965.

[371] K. Schlepckow, J. Wirmer, A. Bachmann, T. Kiefhaber, and H. Schwalbe, "Conserved folding pathways of α-lactalbumin and lysozyme revealed by kinetic CD, fluorescence, NMR, and interrupted refolding experiments," *Journal of Molecular Biology*, vol. 378, no. 3, pp. 686–98, 2008.

[372] R. Huber and W. Bode, "Structural Basis of the Activation, Action and Inhibition of Trypsin," *Accounts of Chemical Research*, vol. 360, no. 4, pp. 489–489, 1979.

[373] P. G. Isenegger and B. G. Davis, "Concepts of Catalysis in Site-Selective Protein Modifications," *Journal of the American Chemical Society*, vol. 141, no. 20, pp. 8005–8013, 2019.

[374] P. Hagel, J. J. Gerding, W. Fieggen, and H. Bloemendal, "Cyanate formation in solutions of urea. I. Calculation of cyanate concentrations at different temperature and pH," *Biochimica et Biophysica Acta*, vol. 243, no. 3, pp. 366–73, 1971.

[375] M. K. Patel, B. Vijayakrishnan, J. R. Koeppe, J. M. Chalker, K. J. Doores, and B. G. Davis, "Analysis of the dispersity in carbohydrate loading of synthetic glycoproteins using maldi-tof mass spectrometry," *Chemical Communications*, vol. 46, no. 48, pp. 9119–9121, 2010.

[376] S. M. Agten, P. E. Dawson, and T. M. Hackeng, "Oxime conjugation in protein chemistry: from carbonyl incorporation to nucleophilic catalysis," *Journal of Peptide Science*, vol. 22, no. 5, pp. 271–279, 2016.

[377] F. Brotzel, Y. C. Chu, and H. Mayr, "Nucleophilicities of primary and secondary amines in water," *Journal of Organic Chemistry*, vol. 72, no. 10, pp. 3679–3688, 2007.

[378] A. B. T. Ghisaidoobe and S. J. Chung, "Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Forster Resonance Energy Transfer Techniques," *International Journal of Molecular Sciences*, vol. 15, no. 12, pp. 22518–22538, 2014.

[379] M. Levitus and S. Ranjit, "Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments," *Quarterly Reviews of Biophysics*, vol. 44, no. 1, pp. 123–151, 2011.

[380] M. You, H. R. Fan, Y. Wang, and W. K. Zhang, "Aldehyde-derivatized indoles as fluorescent probes for hydration environments," *Chemical Physics*, vol. 526, p. 110438, 2019.

[381] J. R. Casasfinet, S. H. Wilson, and R. L. Karpel, "Selective Photochemical Modification by Trichloroethanol of Tryptophan Residues in Proteins with a High Tyrosine-to-Tryptophan Ratio," *Analytical Biochemistry*, vol. 205, no. 1, pp. 27–35, 1992.

[382] R. A. Edwards, G. Jickling, and R. J. Turner, "The light-induced reactions of tryptophan with halocompounds," *Photochemistry and Photobiology*, vol. 75, no. 4, pp. 362–368, 2002.

[383] D. Kazmin, R. A. Edwards, R. J. Turner, E. Larson, and J. Starkey, "Visualization of proteins in acrylamide gels using ultraviolet illumination," *Analytical Biochemistry*, vol. 301, no. 1, pp. 91–96, 2002.

[384] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy - Chapter 16*, pp. 529–575. Boston, MA: Springer US, 2006.

[385] W. Zhang, X. X. Xiang, J. Y. Chen, C. Yang, Y. L. Pan, J. P. Cheng, Q. B. Meng, and X. Li, "Direct C-H difluoromethylation of heterocycles via organic photoredox catalysis," *Nature Communications*, vol. 11, no. 1, 2020.

[386] M. A. Rubinski, S. E. Lopez, and W. R. Dolbier, "Direct access to 2-difluoromethyl indoles via photoredox catalysis," *Journal of Fluorine Chemistry*, vol. 224, pp. 80–88, 2019.

[387] Z. P. A. Wang, Y. Zeng, Y. Kurra, X. Wang, J. M. Tharp, E. C. Vatansever, W. W. Hsu, S. Dai, X. Q. Fang, and W. S. R. Liu, "A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation," *Angewandte Chemie - International Edition*, vol. 56, no. 1, pp. 212–216, 2017.

[388] J. Hofsteenge, D. R. Muller, T. de Beer, A. Loffler, W. J. Richter, and J. F. Vliegenthart, "New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us," *Biochemistry*, vol. 33, no. 46, pp. 13524–30, 1994.

[389] A. Shcherbakova, M. Preller, M. H. Taft, J. Pujols, S. Ventura, B. Tiemann, F. F. R. Buettner, and H. Bakker, "C-mannosylation supports folding and enhances stability of thrombospondin repeats," *eLife*, vol. 8, 2019.

[390] C. E. Murar, F. Thuaud, and J. W. Bode, "KAHA Ligations That Form Aspartyl Aldehyde Residues as Synthetic Handles for Protein Modification and Purification," *Journal of the American Chemical Society*, vol. 136, no. 52, pp. 18140–18148, 2014.

[391] S. Mahesh, V. Adebomi, Z. P. Muneeswaran, and M. Raj, "Bioinspired Nitroalkylation for Selective Protein Modification and Peptide Stapling," *Angewandte Chemie - International Edition*, vol. 59, no. 7, pp. 2793–2801, 2020.

[392] D. Rose-Sperling, M. A. Tran, L. M. Lauth, B. Goretzki, and U. A. Hellmich, "^{19}F NMR as a versatile tool to study membrane protein structure and dynamics," *Biological Chemistry*, vol. 400, no. 10, pp. 1277–1288, 2019.

[393] A. L. P. Lemos, L. Trump, B. Lallemand, P. Pasau, J. Mercier, C. Lemaire, J. C. Monbaliu, C. Genicot, and A. Luxen, "Radical C-H ^{18}F-Difluoromethylation of Heteroarenes with [^{18}F]Difluoromethyl Heteroaryl-Sulfones by Visible Light Photoredox Catalysis," *Catalysts*, vol. 10, no. 275, 2020.

[394] B. Josephson, C. Fehl, P. G. Isenegger, S. Nadal, T. H. Wright, A. W. J. Poh, B. J. Bower, A. M. Giltrap, L. Chen, C. Batchelor-McAuley, G. Roper, O. Arisa, J. B. I. Sap, A. Kawamura, A. J. Baldwin, S. Mohammed, R. G. Compton, V. Gouverneur, and B. G. Davis, "Light-driven post-translational installation of reactive protein side chains," *Nature*, vol. 585, no. 7826, pp. 530–537, 2020.

[395] W. J. Miao, C. F. Ni, Y. C. Zhao, and J. B. Hu, "Radical (2-pyridylsulfonyl)difluoromethylation of terminal alkenes with iododifluoromethyl 2-pyridyl sulfone," *Journal of Fluorine Chemistry*, vol. 167, pp. 231–236, 2014.

[396] G. K. S. Prakash, C. F. Ni, F. Wang, J. B. Hu, and G. A. Olah, "From Difluoromethyl 2-Pyridyl Sulfone to Difluorinated Sulfonates: A Protocol for Nucleophilic Difluoro(sulfonato)methylation," *Angewandte Chemie - International Edition*, vol. 50, no. 11, pp. 2559–2563, 2011.

[397] R. E. Galardy, L. C. Craig, J. D. Jamieson, and M. P. Printz, "Photoaffinity Labeling of Peptide Hormone Binding-Sites," *Journal of Biological Chemistry*, vol. 249, no. 11, pp. 3510–3518, 1974.

[398] J. W. Bode, "Chemical Protein Synthesis with the α-Ketoacid-Hydroxylamine Ligation," *Accounts of Chemical Research*, vol. 50, no. 9, pp. 2104–2115, 2017.

[399] B. Bower, *Novel protein modification strategies*. PhD thesis, University of Oxford, 2018.