RSC Chemical Biology

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue



Cite this: *RSC Chem. Biol.*, 2025, **6**, 1740

# $^{19}$F NMR-tags for peptidyl prolyl conformation analysis

George S. M. Hanson,[a] Faidra Batsaki,[a] Teagan L. Myerscough,[a] Kristin Piché,[b] Ariel Louwrier[b] and Christopher R. Coxon *[a]

Proline *cis*/*trans* isomerism plays an important role in protein folding and mediating protein–protein interactions in short linear interacting motifs within intrinsically disordered protein regions. The slow exchange rate between *cis* and *trans* prolyl bonds provides distinct signals in $^{19}$F NMR analysis of fluorinated peptides, allowing for simple quantification of each population. However, fluorine is not naturally found in proteins but can be introduced using chemical tags. In this study, we evaluate a range of fluorinated cysteine-reactive $^{19}$F NMR tags to assess their ability to react with short, linear proline-containing peptides and accurately report on the equilibrium *cis*/*trans*-Pro populations. Several fluorinated electrophilic tags, including nitrobenzenes, sulfonylpyrimidines, and acrylamides, were found to react chemoselectively and reliably report on the %*cis*-Pro in the model peptide Ac-LPAAC. Other $^{19}$F NMR tags were found to be poor reporters of local proline conformation. Although pentafluoropyridine was non-chemoselective, it still reliably reported on %*cis*-Pro when conjugated *via* cysteine or tyrosine in Ac-LPAAX (X = Cys, Tyr, Lys) peptides. 3,4-Difluoronitrobenzene was found to be compatible with protein tagging, albeit it had modest reactivity and afforded a pair of regioisomeric tagging-products when reacted with a cysteine mutant of α-synuclein. These tools may be valuable for probing *cis*/*trans*-Pro populations in proteins.

Received 9th May 2025,
Accepted 27th August 2025

DOI: 10.1039/d5cb00118h

rsc.li/rsc-chembio



## Introduction

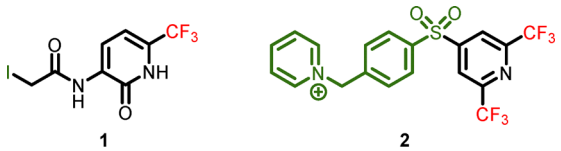
$^{19}$F NMR is a valuable tool for the study of biological processes including protein folding, interactions and aggregation.[1] $^{19}$F NMR exhibits a wide chemical shift window (~300 ppm) and is highly sensitive to its local chemical environment. $^{19}$F is a spin $\frac{1}{2}$ nucleus with similar sensitivity to $^{1}$H. It has 100% natural abundance and is almost entirely absent from nature in the organic form, including being absent in proteins. Therefore, $^{19}$F NMR studies of biological samples should exhibit no background signals – a limitation of $^{1}$H NMR – allowing relatively straightforward analysis in native-like conditions.[2,3]

Fluorinated unnatural amino acids can be readily incorporated into peptides and proteins for $^{19}$F NMR studies using solid-phase peptide synthesis or recombinant protein expression by genetic code expansion. However, the introduction of fluorine-tags by site-specific or chemoselective bioconjugation (fluorine-tagging) broadens the range of fluorine labels available, covers a wider range of chemical shifts, allows tuning of sensitivity and installation of multiple different reporters simultaneously. Fluorine-tagging can also, unlike recombinant expression or protein total synthesis methods, incorporate $^{19}$F NMR reporters into native proteins.

Cysteine is commonly targeted for protein fluorine-tagging, using warheads such as haloacetones,[4] *N*-aryl-2-haloacetamides,[5,6] acrylamides,[7] maleimides,[8] benzylbromides[9,10] and 2,2,2-trifluoroethanethiol.[11] For instance, the aliphatic $CF_3$ tag 3-bromo-1,1,1-trifluoroacetone (BTFA) was used to $^{19}$F label the Leucine transporter (LeuT) at a mutant cysteine residue, revealing four separate $^{19}$F NMR resonances due to specific conformational states.[12] The aromatic $CF_3$ tag 6-(trifluoromethyl)-2-pyridone (**1**) labelled cysteine in human serum albumin, enhancing $^{19}$F NMR chemical shift dispersion compared with BTFA due to the presence of solvent-dependent tautomers.[13] Few fluorine tags directly arylate proteins for $^{19}$F NMR studies and this requires further evaluation. One such example is the aromatic bis-$CF_3$ tag bis(2,6-trifluoromethyl)pyridine (**2**), affording high $^{19}$F NMR signal-to-noise ratio and selective cysteine labelling of streptococcal protein G (GB1) to study conformational changes induced by dimerization and increasing [$Ca^{2+}$].[14] However, it is important to consider that introducing fluorine atoms or fluorine reporters *via* bioconjugation, can lead to perturbation of the native behaviour of a host protein.[15–20] Therefore, it is reasonable to be cautious of the possible effects of side chain tagging on both local and global

[a] *EaStChem School of Chemistry, The University of Edinburgh, Joseph Black Building, David Brewster Road, Edinburgh, EH9 3FJ, UK. E-mail: chris.coxon@ed.ac.uk*
[b] *StressMarq Biosciences Inc., Hillside Avenue, Victoria, British Columbia, V8T 2C1, Canada*

conformational preferences. Therefore, tags should ideally be small in size and avoid introducing significant changes in polarity or charge.




Leaving groups in green; $^{19}$F NMR reporters in red

Proteins are dynamic species that normally require the adoption of correctly folded states to perform their intended roles in biology. However, despite the long-established structure–function paradigm of proteins, up to 20% of eukaryotic proteins are intrinsically disordered (IDPs) or contain intrinsically disordered regions (IDRs),[21] allowing them to adopt a variety of transient conformations to engage with different binding partners, expanding their range of roles.[22–24] Indeed, the majority of proteins exhibit distinct dynamic conformational changes involved in their normal functions. Prolyl bonds are a key facilitator of conformational change and the slow exchange between *cis*-Pro and *trans*-Pro isomers by rotation at the tertiary amide bond is often the rate-limiting step in protein folding.[25] Proline is also often enriched in the sequences of short linear interacting motifs (SLiMs) found in IDRs and plays a role in their flexibility and broad range of binding partners.[26,27] In some cases, one specific conformer has a significantly higher propensity to form an interaction than the other *e.g.* in the interaction between the prolactin receptor and 14–3–3 proteins a *cis*-Pro had affinity three orders of magnitude greater than *trans*-Pro.[28]

Importantly, the *cis*/*trans*-Pro populations in SLiMs are not dictated by protein tertiary structures and are mostly defined by their local sequence.[29,30] We have previously shown that simple 'conformational balance' peptides, which include a mid-sequence proline and distal fluorinated amino acids (*e.g.*, 4-fluorophenylalanine) introduced *via* SPPS, can effectively report on the influence of proximal amino acids on the populations of *cis*-Pro and *trans*-Pro.[31] This was achieved through the integration of discrete $^{19}$F NMR signals owing to the slow rate of exchange between *cis*-Pro and *trans*-Pro conformations under ambient conditions (Fig. 1A). For translation to whole protein systems, new $^{19}$F NMR tags that report on proline *cis–trans* isomerisation are needed. In this work, we evaluate small, fluorinated 'tags' for their reactivity with a model thiol nucleophile (*N*-acetylcysteine) under biomimetic conditions and then apply selected examples to a proline-containing model peptide with a conjugatable cysteine, to reveal how these tags report on the *cis*/*trans*-Pro populations (Fig. 1B). The fluorine tags identified could in future be applied broadly to the fluorine-tagging of dynamic whole proteins, including SLiMs to study their folding and interactions.

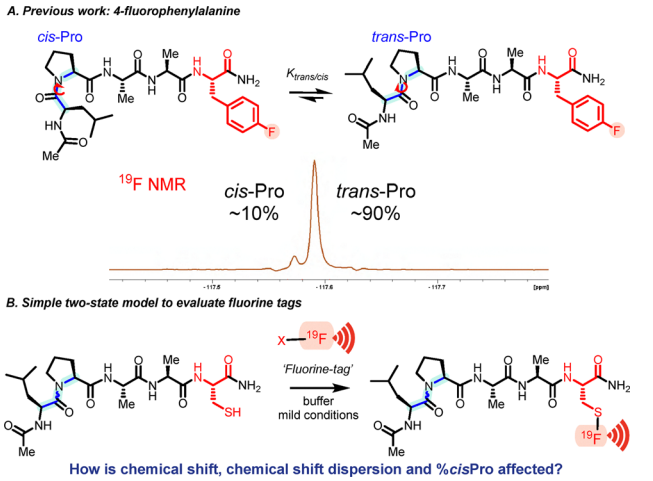



Fig. 1 (A) Fluorinated amino acids can report on *cis*/*trans*-Pro populations in simple conformational balance model peptides by $^{19}$F NMR. (B) By adapting this model to include a conjugatable cysteine, new fluorine-tags can be evaluated as $^{19}$F NMR reporters of conformation.

## Results and discussion

### Reactivity screening of putative fluorinated protein tags

Initially, a small number of putative cysteine-reactive fluorinated reagents were identified as potential protein tags. This comprised a variety of different chemotypes and reactivities, including halo-aromatics (**3–21**; $S_NAr$), thiols (**22**; disulfide formation), benzyl bromides and α-halo ketones (**23–25**; $S_N2$) acrylates and maleimides (**26**, **27**; conjugate-addition) (Fig. 2). These were mostly commercially available reagents with low molecular weight (<300 Da) and *c* Log*P* values (<3.0) to hopefully minimise their impact upon the tagged protein or peptide and to retain water-solubility. Many of the reagents contained aryl-$CF_3$ groups that afford greater signal-to-noise ratio than a single F and are reported to provide greater chemical shift dispersion compared to alkyl-$CF_3$ and reduced chemical shift anisotropy compared to aryl-F.[9] Additionally, three fluorinated acrylamide-based tags (conjugate-acceptors) **28–30** were synthesised by reaction of a fluorinated amine with equimolar acrylic anhydride at room temperature for 0.5 h in acetonitrile without base (**28**, **29**) or by slow addition of $NaHCO_3$ for synthesis of **30** (Scheme 1).

To compare the reactivity of the fluorinated electrophilic reagents, compounds **3–30** (2 eq.) were treated with *N*-acetylcysteine (NAC) (1 eq.) in a solution of water–acetonitrile with DIPEA as a base at 23 °C (Scheme 2), and after 4 h, crude reactions were diluted 10-fold into water before analysis by analytical HPLC. The initial screening ruled out several of the fluorinated small molecules for further evaluation due to poor/no reactivity (summarised in Fig. 2), including difluorobenzamides **3** and **4**, difluorobenzenesulfonamides **5** and **6**, difluoro- and trifluoropyridines **7**, **8**, and **9**, and halobenzotrifluorides **13–15**. The halo(trifluoromethyl)pyridines **10–12** were also, somewhat surprisingly, unreactive despite their electron-deficient nature.

The remaining compounds **16–30** showed varying reactivity towards NAC and the desired conjugate could be identified by

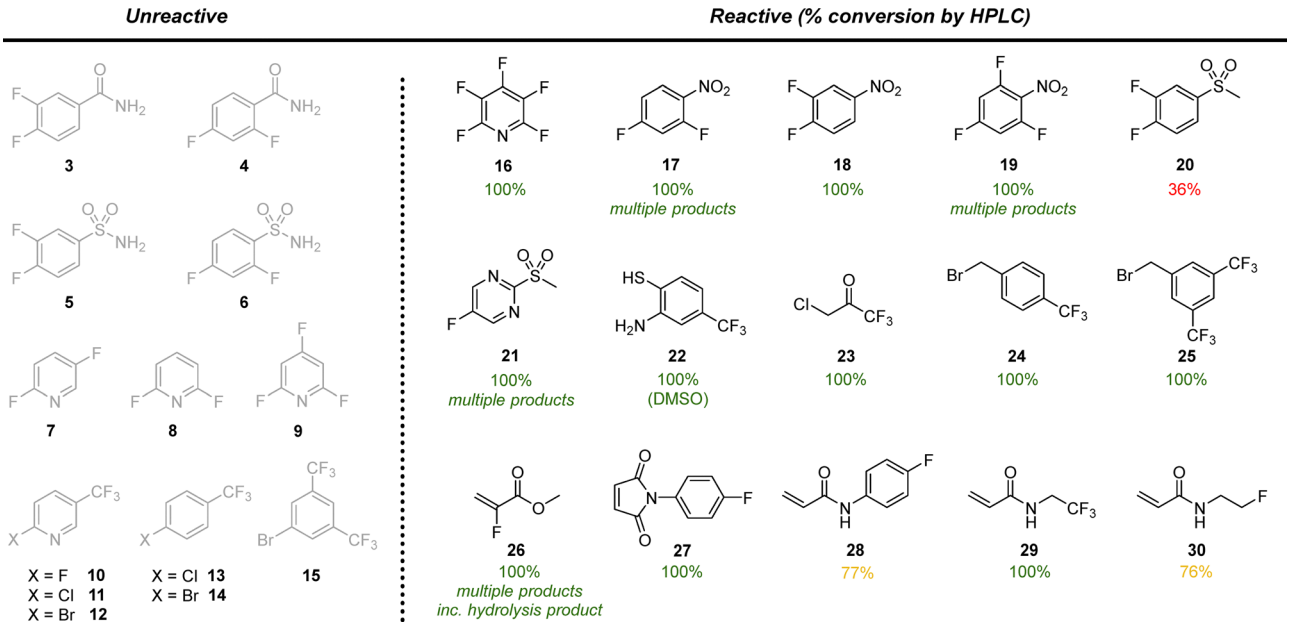
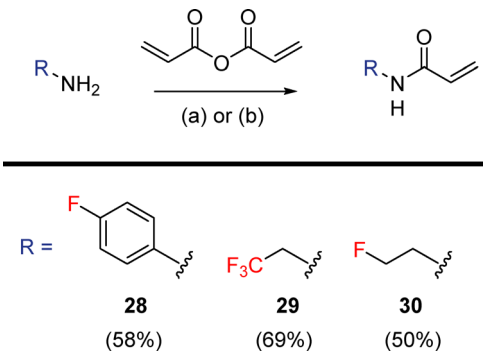
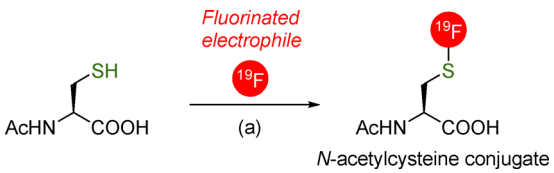



Fig. 2 Reactivity of fluorinated compounds with N-acetyl cysteine (NAC).

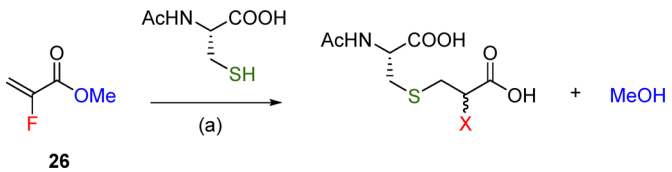



Scheme 1 Synthesis of acrylamides from amines and acrylic anhydride. Reagents and conditions for compounds **28** and **29**: amine (1 eq.), acrylic anhydride (1 eq.), 23 °C, 0.5 h. Compound **30**: amine·HCl (1 eq.), acrylic anhydride (1 eq.), water, increase pH from 2.0–7.0 with $NaHCO_3$.



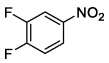
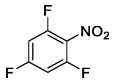
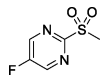
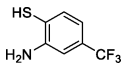
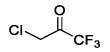
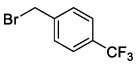
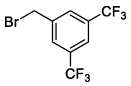
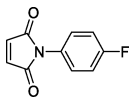
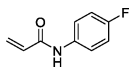
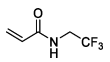

Scheme 2 Reaction screening of fluorine tags. Reagents and conditions: (a) fluorine tag (2 eq.), *N*-acetyl cysteine (1 eq.), DIPEA (5 eq.), water–acetonitrile (1 : 1), 23 °C, 4 h.

LCMS in most cases. Pentafluoropyridine **16** is known to be reactive towards most nucleophilic amino acid sidechains, giving N-, O- and S-centred conjugates,[32] and was found to be highly reactive in our screen. The fluoro-nitrobenzenes **17–19** also showed good reactivity, although **17** and **19** both produced regioisomeric product mixtures. Fluoro-nitrobenzene **18** seemed to undergo selective substitution at the fluorine *para* to the nitro group judging by the single remaining signal in the $^{19}F$ NMR. The difluorophenyl(methyl)sulfone **20** reacted cleanly with NAC, albeit relatively slowly (conversion 36% after 4 h) and was not explored further. The sulfonylpyrimidine **21** is reported to be reactive with cysteine,[33] and was found to react well in our model reaction, albeit with some formation of the unwanted fluorine-substitution product. 3-Chloro-trifluoroacetone **23**, benzylbromides **24** and **25**, *N*-(4-fluorophenyl)maleimide **27** and acrylamides **28–30** also mostly reacted cleanly under these conditions, although the *N*-(4-fluorophenyl)acrylamide **28** and *N*-(2-fluoroethyl)acrymalide **30** were slightly slow (77% and 76% conversion, respectively, after 4 h). The thiophenol **22** reacted with NAC forming a disulfide bond, which was likely aided by the DMSO used as solvent, therefore, this tag may require DMSO to be used as an additive in protein tagging applications. The acrylate **26** underwent the expected conjugate addition with NAC affording a mixture of diastereoisomers. Interestingly, the conjugation product was also observed to hydrolyse during the reaction according to LCMS analysis (see SI), releasing methanol (Scheme 3), unlike the related acrylamides **28–30**. This may represent a useful reaction to conjugate a masked carboxylic acid to a cysteine thiol.




Scheme 3 Reagents and conditions: (a) DIPEA, water–acetonitrile (1 : 1), 23 °C, 4 h.

Table 1 Conjugation of fluorine-tags to conformational balance peptide and reporting of %*cis*-Pro by $^{19}$F NMR

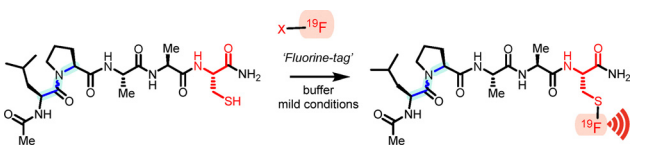
| Fluorine-tag | | Ac-LPAAC conjugate | Conjugate isolated yield (%) | $^{19}$F NMR reported % *cis*-Pro | Dispersion[a] (ppm) |
|---|---|---|---|---|---|
| **16** |  | **31** | 47 | Multiplet (10%)[c] | 0.07[c] |
| **18** |  | **32** | 44 | 10% | 0.06 |
| **19** |  | **33** | 29 | 16% | 0.05 |
| **21** |  | **34** | 32 | 14% | 0.05 |
| **22** |  | **35** | 37 | 10% | 0.04 |
| **23** |  | **36** | 44 | 16% | 0.05 |
| **24** |  | **37** | 24 | Singlet | — |
| **25** |  | **38** | 27 | Singlet | — |
| **26** |  | **39** | 21 | 12%[b] | 0.05[b] |
| **27** |  | **40** | 31 | 24%[d] | 0.07[d] |
| **28** |  | **41** | 39 | 25%[d] | 0.06[d] |
| **29** |  | **42** | 47 | 9%[c] | 0.06[c] |

[a] Chemical shift dispersion measured as the difference in chemical shift for *cis*-Pro and *trans*-Pro resonances. [b] Pair of signals observed. [c] After pure shift $^{19}$F NMR to resolve multiplet. [d] Likely not reporting *cis*-Pro.

## Probing prolyl conformation in two-state peptide models

Having identified several fluorinated electrophiles **16–30** that reacted with NAC under mild conditions, the next objectives were to (i) evaluate how well this translated to the conjugation with a cysteine-containing small peptide model and (ii) test whether each fluorinated tag could report on simple conformational differences by $^{19}$F NMR. The model peptide Ac-LPAA(4-FPhe)-$NH_2$ was previously characterised as a simple two-state 'conformational balance' containing a proline residue affording a mixture of *cis*-Pro and *trans*-Pro conformers (~7–12% *cis*-Pro, Fig. 1A),[29,31] albeit with relatively poor chemical shift dispersion (0.02 ppm) between the *cis*-Pro and *trans*-Pro signals by $^{19}$F NMR. Here, the model peptide was synthesised by standard microwave-assisted Fmoc/$^t$Bu-SPPS and a conjugatable cysteine residue replaced the 4-fluorophenylalanine (4F-Phe) from the earlier study. The cysteine-containing peptide was next treated with a selection of fluorinated electrophiles (Table 1) under the conditions used earlier (3.89 mM peptide, 7.78 mM tag, DIPEA, acetonitrile–water, 23 °C, 4 h) to obtain the conjugates (Scheme 4). The isolated yields of the tagged peptides were mostly moderate-to-low (below 40%), with some **16**, **18**, **23** and **29** affording fair yields (above 40%) (Table 1). The low yields were likely due to product loss during preparative HPLC purification by lyophilization, with reactions being conducted on a small scale (<5 mg crude peptide).

As a benchmark, $^1$H NMR analysis (pH 4.1) of the unmodified cysteine peptide showed the presence of two prolyl bond conformers by NOESY NMR with a relative population of ~11.9% *cis*Pro based on the averaged integration of amide NH resonances (see SI). Assigning $^1$H NMR spectra of large peptides and proteins is challenging due to spectral crowding and signal overlap. Additionally, analysis is often conducted at acidic pH to prevent amide hydrogen exchange. In contrast, $^{19}$F NMR offers distinct advantages, including the ability to study




Scheme 4 General scheme for the fluorine-tagging of a 'conformational balance' model peptide Ac-LPAAC-$NH_2$ for analysis by $^{19}$F NMR. Reagents and conditions: DIPEA, water–acetonitrile (1:1), 23 °C, 4 h.

prolyl bond conformations at neutral pH and a producing more easily interpretable spectrum.

To evaluate the ability of the tags to report on (or influence) prolyl bond *cis–trans* populations, the purified tagged-peptides were dissolved in PBS buffer (pH 7.4, 10% $D_2O$) at a concentration of 1.5 mM and analysed at 23 °C by standard 1D $^{19}F$ NMR (32 scans, 375 MHz). Proton heteronuclear decoupling of $^{19}F$ NMR spectroscopy was used to simplify spectra and improve signal quantification in cases where F–H coupling is possible and would result in multiplets. Tags **18–21**, **26–28** all benefit from decoupling, whereas, in general the $CF_3$ containing tags *e.g.* **22–25** and **29** do not require proton decoupling. Based on our previous study of the X-Pro peptide models,[31] and the above $^1H$ NMR analysis, the Leu-Pro motif was expected to exhibit ~7–12% *cis*-Pro. In most cases, the $^{19}F$ NMR spectra of tagged peptides exhibited the expected pair of singlets representing the *cis*-Pro and *trans*-Pro (see SI).

3,4-Difluoro-nitrobenzene (**18**) again afforded a single regioselective substitution product **32**. By integration of the resulting two NMR signals, the prolyl-bond status was estimated to be ~10% *cis*-Pro (Fig. 3) in good agreement with the $^1H$ NMR analysis of the parent peptide. The thiophenol **22** disulfide conjugate **35** also reported 10% *cis*-Pro, however, **22** did not react without using DMSO as solvent.

The sulfonyl-pyrimidine **21** showed good reactivity towards the peptidyl cysteine and chemoselectivity for sulfone substitution over fluoride. The resulting peptide conjugate **34** displayed the characteristic *cis–trans* NMR signals, albeit the $^{19}F$ NMR sensitivity was relatively poor compared with *e.g.* **18** and displayed low signal to noise ratio due to lower solubility (see SI). Trifluoronitrobenzene **19**, which had shown signs of poor regioselectivity with the initial NAC screen, this time conjugated to the peptide thiol relatively cleanly in the 4-position to give peptide **33**. This afforded two equivalent fluorine environments with double the NMR signal integration compared to **32**, but this tag reported slightly higher %*cis*-Pro content (16%) than expected. The 3-chloro-trifluoroacetone conjugate peptide **36** exhibited a stronger NMR signal with its three-equivalent fluorine atoms, however, **36** again reported slightly higher %*cis*-Pro content (16%) than expected. The fluoroacrylate conjugate **39** again produced a pair of diastereomers (Fig. 3), which was not separated during HPLC purification. The $^{19}F$ NMR analysis revealed that the two diastereoisomers exhibited unique chemical shifts and each of these were split into the characteristic unequal pair of prolyl-isomers (Fig. 3). Despite this, both pairs of signals accurately reported ~12% *cis*-Pro population based on relative peak area. Acrylamide conjugate **42** reported ~9% *cis*-Pro, in line with the anticipated population.

The conjugation of pentafluoropyridine **16** to Ac-LPAAC-$NH_2$ initially afforded a complex $^{19}F$ NMR spectrum (Fig. 4) owing to the F–F scalar coupling of the two different fluorine environments within the tetrafluoropyridyl-conjugate **31**. Thus, it was impossible to directly quantify the populations of prolyl conformers from the NMR spectrum. A simple solution to this was to employ a pure shift broadband homonuclear decoupling $^{19}F$ NMR pulse sequence[34] to remove the fluorine–fluorine scalar coupling and collapse the multiplets centred around approximately −93.5 and −137.8 ppm into pairs of singlets with a chemical shift dispersion of ~0.07 ppm, revealing the presence of *cis*/*trans*-Pro conformers (Fig. 4). The pair of singlets observed for the 2-/5-position fluorine environment observed as the downfield resonance (−93.55 and −93.62 ppm) reported a *cis*-Pro population of 10%, which was consistent with the expected population.

The highly electrophilic **16** is known to also react with the N- and O-nucleophiles of lysine and tyrosine, respectively.[32] Therefore, it was of interest to probe how the $^{19}F$ NMR spectrum (*e.g.* chemical shift and dispersion of the signals) is affected by changing the residue through which the tag was conjugated *i.e.* S, N or O nucleophiles and the nature of the sidechain (*e.g.*

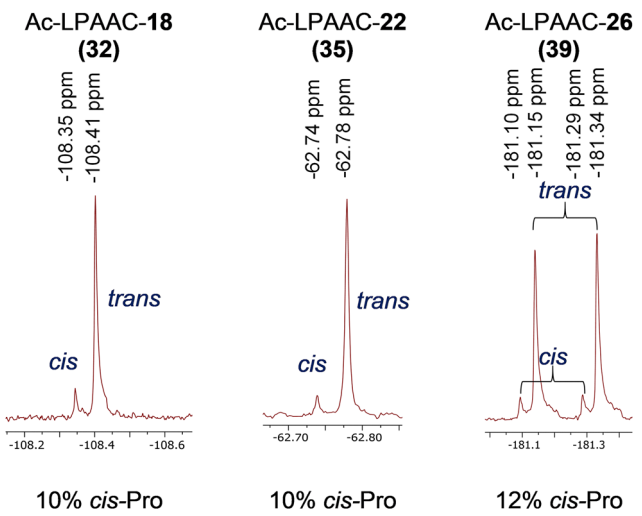



Fig. 3 Example $^{19}F$ NMR spectra of tagged Ac-LPAAC (not to scale).

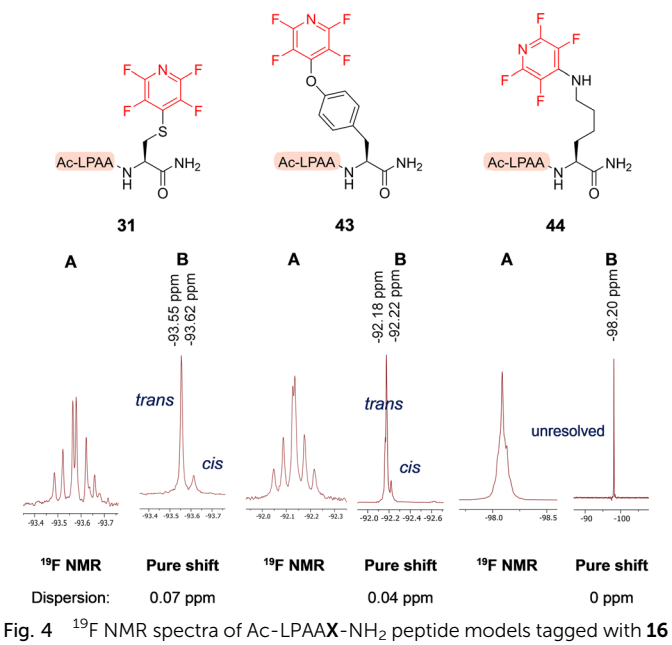



Fig. 4 $^{19}F$ NMR spectra of Ac-LPAA**X**-$NH_2$ peptide models tagged with **16** (not to scale). (A) with F–F coupling and (B) after pure shift F–F decoupling.

length, flexibility). Therefore, two new model peptides Ac-LPAAY and Ac-LPAAK were synthesised based on the same model peptide but replacing cysteine with lysine and tyrosine, respectively. The peptides were conjugated with **16** and analysed using $^{19}$F NMR. In each case, the $^{19}$F NMR spectrum displayed a similar pair of overlapping multiplets or broad unresolved resonances, and therefore, a pure shift pulse sequence was employed. The Tyr O-conjugate **43** exhibited a pair of downfield inequivalent singlets, that resembled those seen in the cysteine peptide (9% *cis*-Pro) but with different chemical shifts (−92.18 and −92.22 ppm) (Fig. 4). Therefore, **16** may have use in tagging and reporting simultaneously at different protein sites. However, for the Lys N-conjugate **44**, the indistinguishable multiplet (−98.20 ppm) was replaced with an unresolved singlet after applying pure shift and it was not possible to estimate the %*cis*-Pro population. This suggests that fluorine-tagging of proteins with **16** through cysteine and tyrosine provides higher sensitivity to local structural changes. Whilst, tagging through the more flexible lysine side chain may be better suited for observing global protein conformational changes without being masked by local changes.

In general, there were only small differences in chemical shift dispersion observed between *cis*-Pro and *trans*-Pro resonances for each of the different fluorine tags despite their distinct chemotypes. Nonetheless, there was some evidence of a slight increase in chemical shift dispersion for *N*-(4-fluorophenyl)maleimide conjugate **40** (0.07 ppm) and tetrafluoropyridine cysteine conjugate **31** (0.07 ppm), which may be explained by the fluorine reporter being directly attached to the phenyl ring, making it more sensitive to the polarization of the aromatic electron cloud.[9] Indeed, the disulfide aryl-$CF_3$ conjugate **35** afforded slightly smaller dispersion (0.04 ppm). Surprisingly, the trifluoromethylbenzyl-conjugates **37** and **38** – previously reported in protein studies[9,10] – afforded only singlets by $^{19}$F NMR for tagged peptides and did not report distinct signals for prolyl conformers. Conversely, the alkyl-$CF_3$ conjugates **36** (trifluoromethylketone – also used in protein studies;[4,12] 0.05 ppm) and **42** (acrylamide; 0.06 ppm) afforded relatively good dispersion. This is despite contrary evidence that aryl-$CF_3$ tags exhibit improved chemical shift sensitivity over alkyl-$CF_3$ groups.[9] It was also of note that the nature of the side chain conjugated affected the chemical shift dispersion. Tetrafluoropyridine tyrosine O-conjugate **43** exhibited smaller chemical shift dispersion (0.04 ppm) compared with the cysteine S-conjugate **31** (0.07 ppm), whilst there was no dispersion observed for the corresponding lysine N-conjugate. This may, however, be a function of distance from the proline or increased flexibility in the amino acid side chain.

Some tags proved to be of little value for probing proline bond conformations, for a variety of reasons (some noted above). For acrylamide peptide-conjugate **41** the %*cis*-Pro was abnormally high, around 25%, which was initially ascribed to the conjugate actually reporting the more proximal amide bond *cis*/*trans* populations.[35] However, this was unexpected due to the closely related acrylamide **42** displaying ∼9% *cis*-Pro and the reported fluorine tag, 2-bromo-*N*-(4-(trifluoromethyl)phenyl)-acetamide (BTFMA) consistently proving to be useful for probing global protein conformational changes.[9,36] The reason for this discrepancy is unresolved. Maleimide-tagged **40** also reported an unlikely high %*cis*-Pro content (∼25%), perhaps due to the generation of two diastereomeric products from the conjugate addition. Therefore, these tags were deemed to be unreliable reporters of local %*cis*-Pro conformation.

### Tagging reaction kinetics and chemoselectivity in aqueous buffer

To be of value to biological studies, the $^{19}$F NMR tags should also be reactive with specific protein side chains under mild physiological conditions. To compare different conjugation chemistries, nitrobenzene **18**, sulfonylpyrimidine **21** chloroacetone **23** and acrylamide **29** – that were all reactive towards cysteine under semi-aqueous conditions – were mixed with equimolar NAC in Tris buffer (20 mM, pH 7.4) at 37 °C (0 min timepoint). The conjugation reaction was started by the addition of DIPEA (2 eq.) and the consumption of NAC was measured by analytical RP-HPLC peak area after quenching samples after 10, 20 and 30 minutes. Each of the compounds reacted rapidly with the thiol and achieved >50% conversion in <10 minutes (Fig. 5). Acrylamide **29**, as well as chloroacetone **23** and sulfonylpyrimidine **21**, all afforded complete consumption of NAC in this 10-minute window. The nitrobenzene **18** was somewhat slower, disagreeing with earlier experiments using 1:1 water-acetonitrile, affording ∼80% consumption after 30 minutes. Nevertheless, this could still be useful for protein bioconjugation.

To rule out any potential off-target conjugation, **18**, **21**, **23** and **29** were treated with *N*-acetyl lysine (NAK) and *N*-acetyl tyrosine (NAY) under the same conditions as above for four hours with periodic monitoring by HPLC (Table 2). Compounds

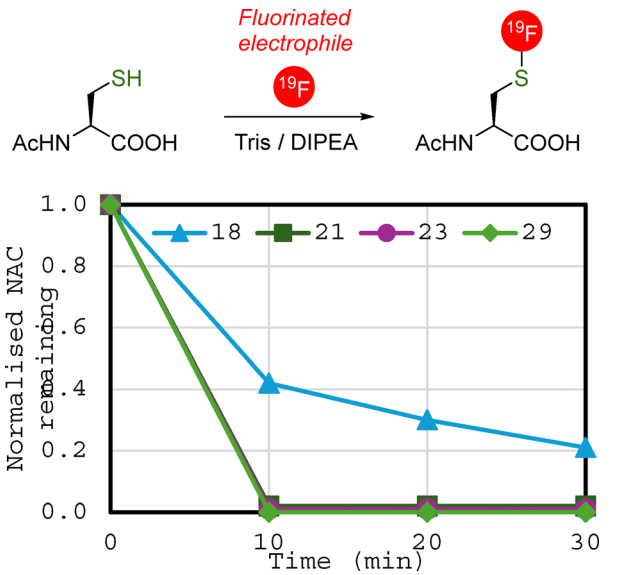
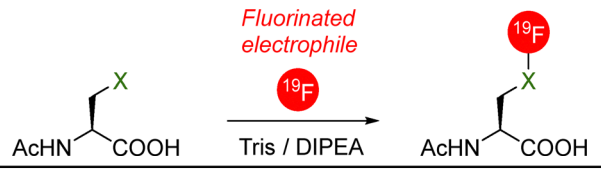
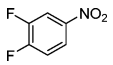
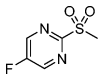
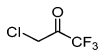
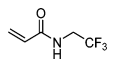



Fig. 5 HPLC reaction kinetics for the potential fluorine-tags **18**, **21**, **23** and **29** with NAC in aqueous Tris buffer at 37 °C.

Table 2 Selectivity screening of compounds with alternative protein nucleophiles NAK and NAY




| Tag | | Consumption of nucleophile (%) NAK X = | NAY X = |
|---|---|---|---|
| **18** |  | 0 | 0 |
| **21** |  | 0 | 0 |
| **23** |  | 54 | 22 |
| **29** |  | 0 | 0 |

**18**, **21** and **29** were unreactive towards either NAK or NAY and are so far unreported in the literature as cysteine-selective $^{19}$F NMR tags. However, chloroacetone **23**, just like **16**, reacted with both NAK and NAY and was non-selective for cysteine. This is despite the closely related bromoacetone being commonly used in the literature for cysteine-tagging.[4]

### Fluorine-tagging an intrinsically-disordered protein

Finally, we aimed to evaluate whether the newly identified tags can also be used to label a protein. α-Synuclein is an intrinsically-disordered protein (IDP) that plays a role in Parkinson's disease by forming misfolded protein aggregates. A recombinant cysteine mutant of α-synuclein (A90C) has been used previously to study aggregate formation by tagging with fluorescent labels for FRET studies.[37,38] Despite its slower kinetics of conjugation with NAC compared with pyrimidine **21**, 3,4-difluoro-nitrobenzene (**18**) was selected as an example tag for direct cysteine arylation at Cys-90 (Fig. 6) because it had afforded a relatively high conjugation yield (44%), reported the %*cis*Pro accurately and exhibited greater chemical shift dispersion than **21** with baseline-resolution in the $^{19}$F NMR of the peptide model.

The tagging reaction of protein with **18** was found to be more challenging than expected. Several different reaction conditions were evaluated, including changing the tag concentration, buffer type, addition of organic base, temperature and reaction time (see SI). Tris buffer (adjusted to pH 8.6) was initially trialled based on earlier kinetics experiments, however, no tagged protein as observed by mass spectrometry after 18 h at 4 °C with 50 eq. of tag; whilst increasing the temperature to 23 °C was still unproductive. The buffer seemed to play a significant role and the reaction of protein (0.34 mM) with **18** (17 mM) in HEPES buffer adjusted to pH 8.6 afforded

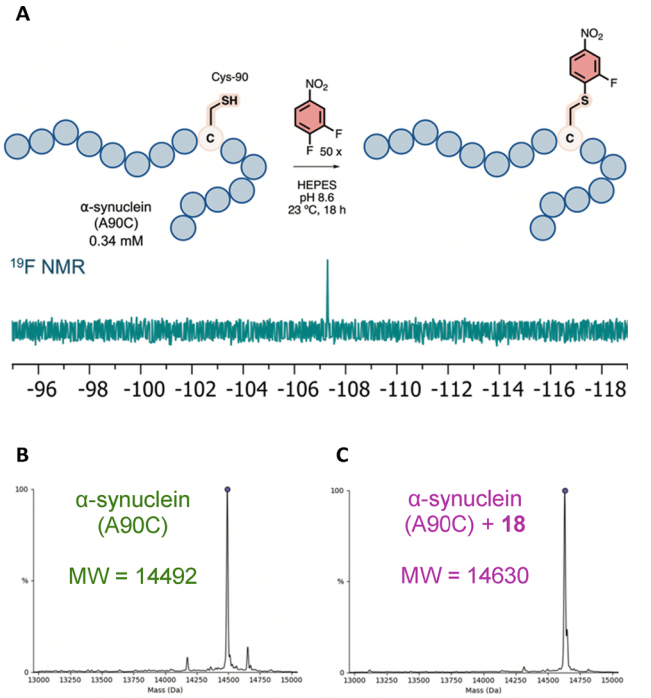



Fig. 6 α-synuclein tagging with **18**. (A) $^{19}$F NMR of purified A90C α-synuclein tagged with **18**. (B) MS of untagged purified A90C α-synuclein; (C) MS of purified A90C α-synuclein tagged with **18**.

completely tagged protein after incubation for 18 h at 23 °C, as confirmed by mass spectrometry (MW = 14 630 Da, Fig. 6B and C). Given the earlier use of DIPEA in the kinetics experiments, we considered that adding the organic base (50 eq.) might allow us to obtain a faster tagging reaction by ensuring full ionisation of the Cys-90 thiolate. Unfortunately, despite also affording tagged product, this led to formation of an additional dehydroalanine product (MW = 14458) resulting from the elimination of the tag from cysteine (see SI). Despite this, it was found that we could obtain tagged protein with only 5 eq. of tag when adding DIPEA. $^{19}$F NMR analysis of the purified and buffer exchanged tagged protein (0.3 mM peptide, PBS buffer pH 7.4, 10% $D_2O$, 23 °C) exhibited a relatively sharp singlet at ∼ −107.29 ppm (Fig. 6). This indicated that the conjugation was regioselective and that a fluorinated protein signal could easily be observed, providing a tool for future studies of protein aggregation by $^{19}$F NMR.

## Conclusions

This work has identified several previously unreported $^{19}$F NMR tags. Some of these exhibit well-dispersed *cis*/*trans*-Pro signals and allow accurate quantification of prolyl bond conformational status from at least three residues away in small model peptides. In most cases, the tags selectively conjugated to cysteine in unprotected peptides with rapid conjugation kinetics (<30 min) under biomimetic conditions. It was notable that when pentafluoropyridine (**16**) was used to tag cysteine, lysine and tyrosine, each of the S-, N- and O-conjugates afforded

distinct chemical shifts and different chemical shift dispersions yet still reported the same %*cis*-Pro (where observable) after applying a broadband homonuclear decoupling pure shift sequence. Therefore, **16** may be useful for tagging different side chains simultaneously to study wider protein conformational changes. It was also evident that some tags either inaccurately reported or perhaps affected the prolyl bond conformational preferences.

In general, most of the tags afforded broadly similar peak widths and chemical shift dispersions between *cis*-Pro and *trans*-Pro resonances, ranging from 0.04 – 0.07 ppm. Tags **16** (after pure shift), **18**, **22**, **28** and **29** all afforded baseline-resolved signals and outperformed previously reported tags including trifluoromethylbenzyl groups **24** and **25** with respect to chemical shift dispersion in this prolyl *cis*/*trans* model. Cysteine-tagging also afforded greater dispersion than our earlier reported fluorinated amino acid 4-fluorophenylalanine (0.02 ppm) in the same position.[31] However, in these peptides the chemical shift dispersion was more significantly affected by the nature of the amino acids proximal to proline (ranged from 0.01 to 0.17 ppm) than we have observed for the different tags reported here.[31]

3,4-Difluoronitrobenzene **18** was found to be compatible with protein tagging, requiring a slightly basic pH of 8.6 and room temperature to undergo nucleophilic aromatic substitution, whereas Michael acceptors such as maleimides are able to react rapidly at neutral pH and at low temperatures. Moreover, the addition of an organic base was detrimental, and led to the elimination of the tag, forming a dehydroalanine. Other tags studied may have been more reactive under milder conditions but were not tested. Future work will explore the specific structural features of $^{19}$F NMR tags that that affect chemical shift dispersion and accurate conformation reporting in SLiMs and IDPs.

## Author contributions

GSMH conducted all synthetic and characterisation studies reported here. TM performed NMR analysis. FB performed protein tagging and characterisation work. KP and AL performed protein expression and purification. CRC and GSMH designed the experiments and wrote the manuscript together.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

All data supporting this article have been included as part of the SI. This includes methods and full spectral characterisation. See DOI: **https://doi.org/10.1039/d5cb00118h**.

## Acknowledgements

This research was funded in part, by the Biotechnology and Biological Sciences Research Council [BB/X019756/1] and The University of Edinburgh BBSRC Impact Acceleration Account [BBSRC IAA PIII141]. The University of Edinburgh provided PhD scholarship funding for GSMH, FB and TM. For the purpose of open access, the author has applied a creative commons attribution (CC BY) licence to any author accepted manuscript version arising.

## Notes and references

1 G. S. Hanson and C. R. Coxon, *ChemBioChem*, 2024, **25**, e202400195.
2 D. Gimenez, A. Phelan, C. D. Murphy and S. L. Cobb, *Beilstein J. Org. Chem.*, 2021, **17**, 293–318.
3 H. Chen, S. Viel, F. Ziarelli and L. Peng, *Chem. Soc. Rev.*, 2013, **42**, 7971–7982.
4 U. A. Hellmich, N. Pfleger and C. Glaubitz, *Photochem. Photobiol.*, 2009, **85**, 535–539.
5 A. Manglik, T. H. Kim, M. Masureel, C. Altenbach, Z. Yang, D. Hilger, M. T. Lerch, T. S. Kobilka, F. S. Thian, W. L. Hubbell, R. S. Prosser and B. K. Kobilka, *Cell*, 2015, **161**, 1101–1111.
6 H. R. Kalbitzer, G. Rohr, E. Nowak, R. S. Goody, W. Kuhn and E. Zimmermann, *NMR Biomed.*, 1992, **5**, 347–350.
7 H. Hiscocks, A. Ung and G. Pascali, *AppliedChem*, 2023, **3**, 256–278.
8 J. M. Edwards, J. P. Derrick, C. F. van der Walle and A. P. Golovanov, *Mol. Pharm.*, 2018, **15**, 2785–2796.
9 L. Ye, S. T. Larda, Y. F. Frank Li, A. Manglik and R. S. Prosser, *J. Biomol. NMR*, 2015, **62**, 97–103.
10 M. Somlyay, K. Ledolter, M. Kitzler, G. Sandford, S. L. Cobb and R. Konrat, *ChemBioChem*, 2020, **21**, 696–701.
11 R. Horst, J. J. Liu, R. C. Stevens and K. Wüthrich, *Angew. Chem., Int. Ed.*, 2013, **52**, 10762–10765.
12 A. Daminato, C. J. Loland and E. J. Cabrita, *J. Neurochem.*, 2025, **169**, e16278.
13 G. A. Frere, A. Hasabnis, C. B. Francisco, M. Suleiman, O. Alimowska, R. Rahmatullah, J. Gould, C. Y. C. Su, O. Voznyy, P. T. Gunning, E. A. Basso and R. S. Prosser, *J. Am. Chem. Soc.*, 2024, **146**, 3052–3064.
14 Z. Chai, Q. Wu, K. Cheng, X. Liu, Z. Zhang, L. Jiang, Z. Zhou, M. Liu and C. Li, *Angew. Chem., Int. Ed.*, 2023, **62**, e202300318.
15 A. A. Berger, J. S. Völler, N. Budisa and B. Koksch, *Acc. Chem. Res.*, 2017, **50**, 2093–2103.
16 S. A. Miles, J. A. Nillama and L. Hunter, *Molecules*, 2023, **28**, 6192.
17 E. S. Eberhardt, N. Panasik and R. T. Raines, *J. Am. Chem. Soc.*, 1996, **118**, 12261–12266.
18 H. Iwai, A. Lingel and A. Pluckthun, *J. Biol. Chem.*, 2001, **276**, 16548–16554.
19 B. C. Buer, J. L. Meagher, J. A. Stuckey and E. N. G. Marsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4810–4815.

20 M. A. Cejas, W. A. Kinney, C. Chen, J. G. Vinter, H. R. Almond, K. M. Balss, C. A. Maryanoff, U. Schmidt, M. Breslav, A. Mahan, E. Lacy and B. E. Maryanoff, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8513–8518.
21 B. Xue, A. K. Dunker and V. N. Uversky, *J. Biomol. Struct. Dyn.*, 2012, **30**, 137–149.
22 C. J. Oldfield and A. K. Dunker, *Annu. Rev. Biochem.*, 2014, **83**, 553–584.
23 H. J. Dyson and P. E. Wright, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 197–208.
24 M. M. Babu, *Biochem. Soc. Trans.*, 2016, **44**, 1185–1200.
25 C. Dugave and L. Demange, *Chem. Rev.*, 2003, **103**, 2475–2532.
26 V. N. Uversky, *Intrinsically Disord. Proteins*, 2013, **1**, e24684.
27 B. Mateos, C. Conrad-Billroth, M. Schiavina, A. Beier, G. Kontaxis, R. Konrat, I. C. Felli and R. Pierattelli, *J. Mol. Biol.*, 2020, **432**, 3093–3111.
28 F. F. Theisen, A. Prestel, N. L. Jacobsen, O. K. Nyhegn-Eriksen, J. G. Olsen and B. B. Kragelund, *J. Am. Chem. Soc.*, 2025, **147**, 5714–5724.
29 U. Reimer, G. Scherer, M. Drewello, S. Kruber, M. Schutkowski and G. Fischer, *J. Mol. Biol.*, 1998, **279**, 449–460.
30 F. Sebák, J. Szolomájer, N. Papp, G. K. Tóth and A. Bodor, *Front. Biosci.-Landmark*, 2023, **28**, 127.
31 P. M. Killoran, G. S. Hanson, S. J. Verhoork, M. Smith, D. Del Gobbo, L. Y. Lian and C. R. Coxon, *Chem. – Eur. J.*, 2023, **29**, e202203017.
32 D. Gimenez, C. A. Mooney, A. Dose, G. Sandford, C. R. Coxon and S. L. Cobb, *Org. Biomol. Chem.*, 2017, **15**, 4086–4095.
33 M. R. Bauer, A. C. Joerger and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E5271–E5280.
34 L. Castañar, P. Nolis, A. Virgili and T. Parella, *Chem. – Eur. J.*, 2013, **19**, 17283–17286.
35 R. Quintanilla-Licea, J. F. Colunga-Valladares, A. Caballero-Quintero, C. Rodríguez-Padilla, R. Tamez-Guerra, R. Gómez-Flores and N. Waksman, *Molecules*, 2002, **7**, 662–673.
36 A. Manglik, T. H. Kim, M. Masureel, C. Altenbach, Z. Yang, D. Hilger, M. T. Lerch, T. S. Kobilka, F. S. Thian, W. L. Hubbell and R. S. Prosser, *Cell*, 2015, **161**, 1101–1111.
37 M. H. Horrocks, L. Tosatto, A. J. Dear, G. A. Garcia, M. Iljina, N. Cremades, M. Dalla Serra, T. P. Knowles, C. M. Dobson and D. Klenerman, *Anal. Chem.*, 2015, **87**, 8818–8826.
38 M. L. Choi, A. Chappard, B. P. Singh, C. Maclachlan, M. Rodrigues, E. I. Fedotova, A. V. Berezhnov, S. De, C. J. Peddie, D. Athauda, G. S. Virdi, W. Zhang, J. R. Evans, A. I. Wernick, Z. S. Zanjani, P. R. Angelova, N. Esteras, A. Y. Vinokurov, K. Morris, K. Jeacock, L. Tosatto, D. Little, P. Gissen, D. J. Clarke, T. Kunath, L. Collinson, D. Klenerman, A. Y. Abramov, M. H. Horrocks and S. Gandhi, *Nat. Neurosci.*, 2022, **25**, 1134–1148.