Site-selective protein modification via peptide-directed proximity catalysis

Laetitia Raynal,[a,b†] Joe Nabarro,[a,b†], Lisa M. Miller,[b,c] Adam A. Dowle,[d] Sophie L. Moul,[a,b] Steven D. Johnson,[b,c] Martin A. Fascione,[a,b] and Christopher D. Spicer*[a,b]

[†]Authors contributed equally
[a]Department of Chemistry, University of York, Heslington, YO10 5DD, UK
[b]York Biomedical Research Institute, University of York, Heslington, YO10 5DD, UK
[c]School of Physics, Engineering, and Technology, University of York, Heslington, YO10 5DD, UK
[d]Bioscience Technology Facility, Department of Biology, University of York, Heslington, YO10 5DD, UK

* chris.spicer@york.ac.uk

ABSTRACT: Proximity catalysis exploits ligand-binding for localised, catalytic protein modification. In this work, we introduce catalyst-functionalised peptides as versatile ligands for this approach. Through the functionalisation of target-binding peptides with pyridinium oximes catalysts, we show that model proteins can be site-selectively modified with a variety of *N*-acyl-*N*-alkylsulfonamide reagents, to introduce common functionalities including fluorophores and affinity handles to the protein surface. Critically, we show that simple changes to the peptide-catalyst structure, moving the pyridinium oxime from N- to C-terminus, alter the site of modification. This opens up possibilities to develop peptide libraries for a particular target protein, and subsequently tuning the modification site for a given application.

Introduction

Site-selective chemical modification remains a crucial challenge in the development of protein-based technologies and tools. *Chemo*-selective approaches to achieve single-site labelling typically rely on the introduction of a uniquely reactive amino acid, most commonly a solvent-exposed cysteine or an unnatural reactive handle via codon reassignment.[1,2] Though these methods are powerful, they are limited by the requirement for prior genetic engineering of the target protein, and are therefore poorly suited to proteins isolated from natural sources, or of eukaryotic origin where recombinant expression may be challenging. As a result, there has been growing interest in the development of *regio*-selective approaches, that can specifically target an amino acid on a protein surface, even in the presence of other residues bearing the same functional groups.[3]

Of these regio-selective approaches, proximity-mediated strategies have proved most powerful to date. In these strategies, a moderately reactive species is brought into close proximity with a protein surface via some form of ligand binding, creating a pseudo-intramolecular environment that enables labelling reactions that would not otherwise take place without this effective concentration.[4]

In most cases, small molecule ligands are used to mediate protein binding, and as a result modification is commonly in proximity to the protein active site.[5] More recently, peptide-based ligands have found utility in proximity labelling[6–9] providing three major advantages: i) with ongoing development in screening technologies such as phage and mRNA display, peptides can be 'evolved' to bind to most target proteins;[10] ii) the lack of bias in these screening processes provides opportunities to identify peptide ligands that bind distal to the active site of a target protein, with less chance of adversely affecting protein activity; and iii) solid-phase synthesis streamlines the synthesis and functionalization of peptides, reducing the challenges of ligand development. Peptides therefore provide a potential route to highly generalisable and translatable protein modification strategies, which can be widely applied to proteins of diverse origin and function.

In this paper, we report the first example of peptide-based proximity *catalysis*, a powerful extension of

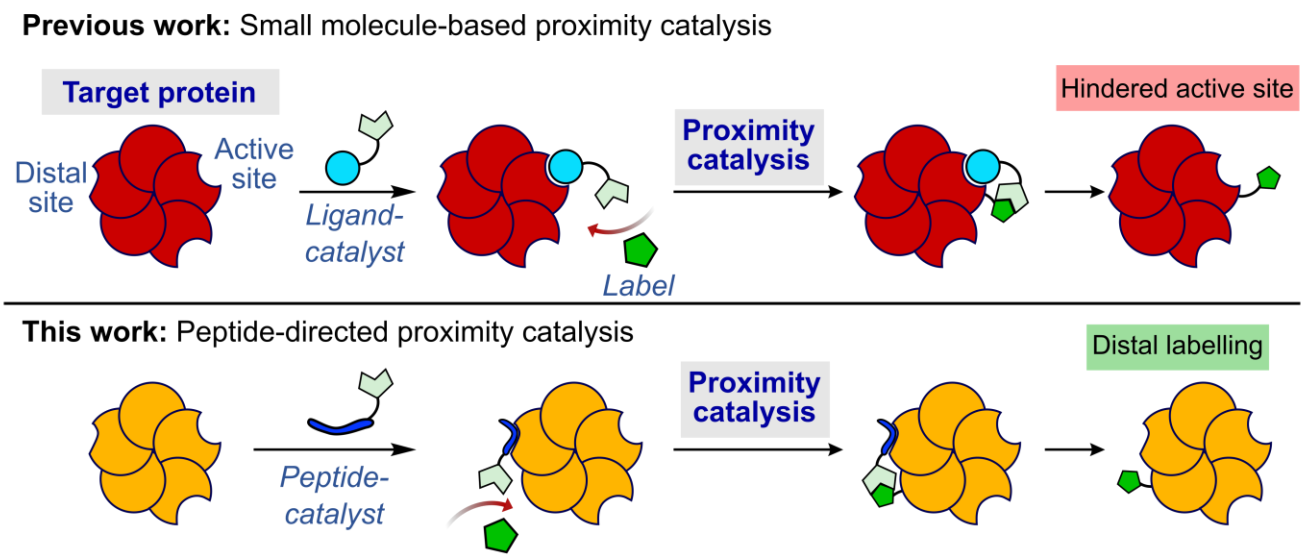


Figure 1: Overview of the use of peptide-directed proximity catalysis to mediate site-selective protein labelling distal to the active site. By comparison, traditional small-molecule ligands most commonly bind and induce labelling of the active site, potentially interfering with protein activity.

traditional proximity modifications that decouples the processes of ligand binding and subsequent protein labelling (***Fig. 1***).[6,11,12] In doing so, it enables the introduction of a diverse range of labels using a single ligand-catalyst species. We demonstrate that different sites on the protein surface can be targeted by tuning the structure of the peptide-catalyst, and that this versatile technology can be applied to a number of different protein targets. We anticipate that our peptide-directed proximity catalysis approach will provide a valuable addition to the toolbox of reactions for site-specific protein modification, particularly for targets that cannot be genetically engineered.

Results and discussion

Our initial project design was inspired by the work of the Hamachi group, who over the past 15 years have pioneered the development of both ligand-directed chemistries and proximity catalysis.[4,13] Of the catalyst systems reported, we were particularly attracted to the use of pyridinium oximes, which can catalyse protein acylation with *N*-acyl-*N*-alkylsulfonamide (NASA) reagents in a proximity-dependent manner.[14] NASAs have been shown to possess improved stability relative to alternative reagents for protein acylation, such as thioesters, while pyridinium oxime catalysts provide faster protein labelling than alternatives such as dimethylaminopyridine- or rhodium-catalysts.[4]

We envisaged introducing pyridinium oximes at strategic positions within our target-binding peptides during their synthesis, which could then be used to modify the target protein with a diverse set of NASA-functionalised labels. As an initial target on which to develop our approach, we chose human insulin. Although insulin can be expressed recombinantly and genetic engineering is therefore plausible, its small size made it an ideal model. A number of peptide sequences have previously been reported to bind insulin.[15–17] Of these, we chose to focus on the peptide RGFFYT (**P1**). This sequence is actually derived from insulin itself, playing a role in insulin self-assembly and oligomerization, and has been computationally predicted to bind parallel to the insulin B-chain.[17] We hypothesised that introduction of a catalyst at either end of the peptide would lead to significant differences in preferred site of insulin labelling.

To predict whether the insulin-binding capability of **P1** would be maintained after the introduction of a pyridinium oxime catalyst, peptides bearing either an N- or C-terminal cysteine (**C-P1** and **P1-C** respectively) were first synthesised, and tethered to a gold sensor for quartz crystal microbalance with dissipation monitoring (QCM-D) analysis. Upon addition of insulin, a characteristic binding response was observed for both **C-P1** and **P1-C**, suggesting the introduction of the pyridinium oxime catalyst at either terminus would be tolerated. From these experiments, dissociation constants (K_ds) of 1.5 μM and 0.8 μM were calculated following N-terminal and C-terminal tethering respectively (Supplementary information, *Fig. S1 and S2*).

Peptides were subsequently synthesized bearing either a C-terminal propargylglycine or N-terminal 5-pentynoic acid (**P1-alkyne** and **alkyne-P1** respectively, ***Fig. 2a***). *p*-Nitrophenyl NASA-reagents were also synthesized bearing hexanoic acid (**3a**), biotin (**3b**), nitrobenzoxadiazole (NBD, **3c**), and Cy5 (**3d**) acyl

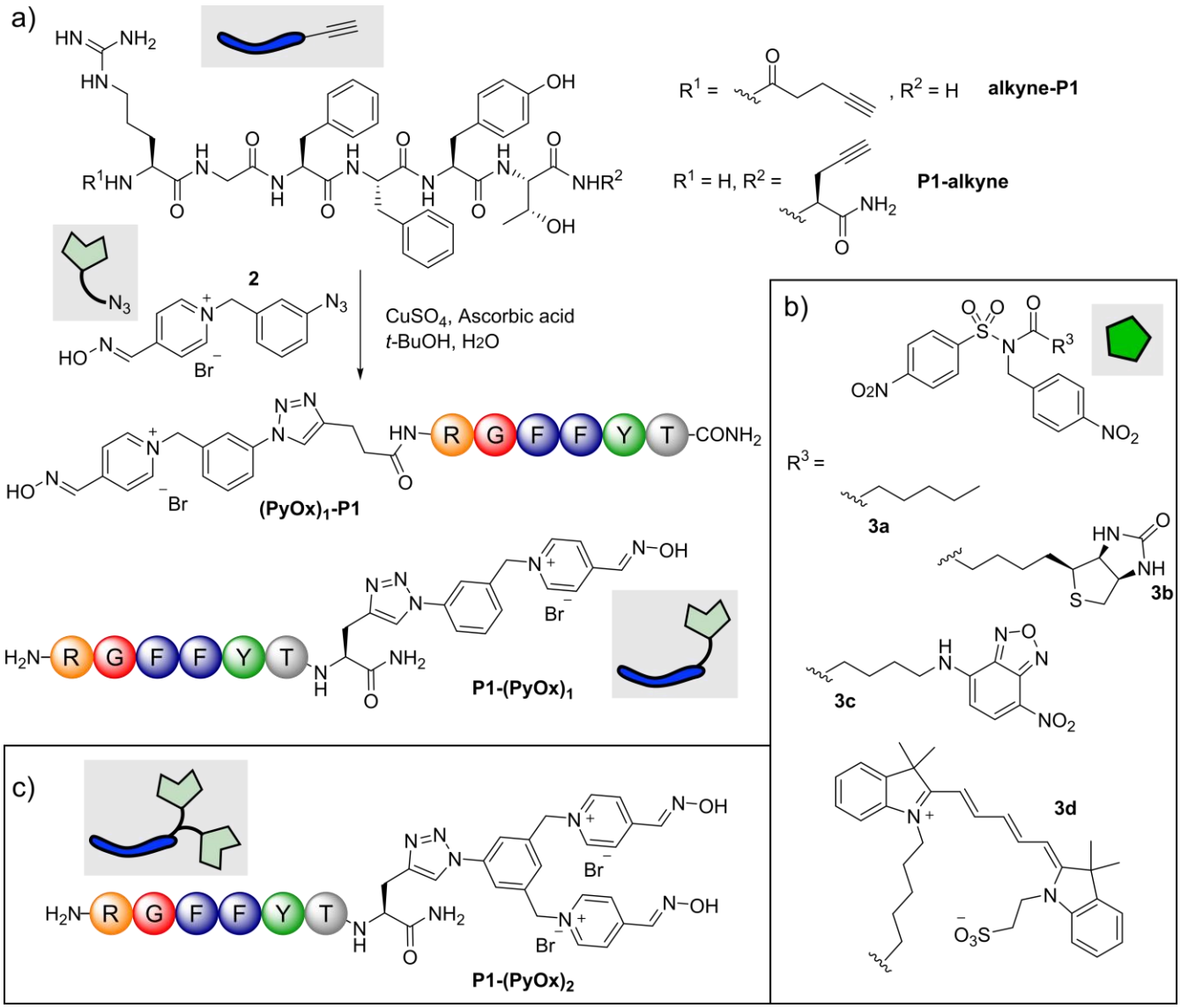
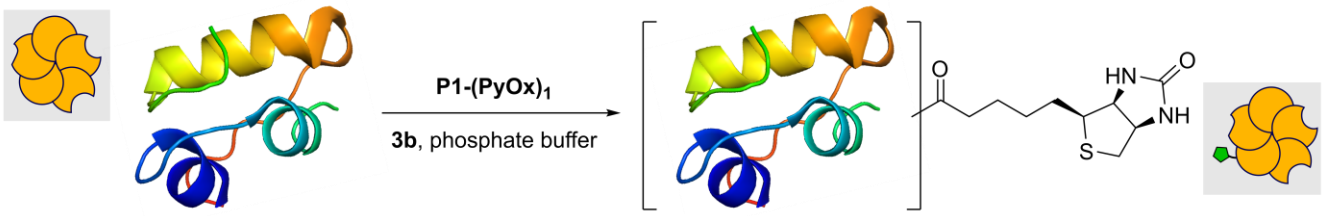


Figure 2: a) Synthesis of pyridinium oxime-functionalised analogues of peptide **P1** via copper-catalysed azide-alkyne cycloaddition; b) *N*-acyl-*N*-alkylsulfonamide (NASA) reagents synthesised and used in this work; c) Bis-pyridinium oxime peptide **P1-(PyOx)$_2$**.

groups, as model labels for subsequent protein modification (***Fig. 2b***).[14]

With these reagents in hand, we went on to attempt peptide-directed proximity catalysis on insulin. Initial experiments were performed at an insulin concentration of 50 μM in pH 7.4 phosphate buffer, which was above the K_d calculated for peptide binding (***Fig. 3***). Varying equivalents of pyridinium oxime peptide **P1-(PyOx)$_1$** (0, 1.25, 2.5, 5, or 10 equiv. w.r.t. protein) were then added, followed by addition of biotin-NASA **3b** (20 equiv. w.r.t. protein). Loadings of the NASA were limited by the poor solubility of **3b** in aqueous media, requiring pre-solubilisation in DMSO, which was subsequently diluted in the reaction mixture to a final concentration of ~1% v/v. After 2 h at room temperature (22 °C), reactions were analysed by in-tact protein liquid-chromatography mass spectrometry (LC-MS). Pleasingly, in the presence of both **P1-(PyOx)$_1$** and **3b** a mass corresponding to acylation of insulin with biotin was observed (*Entries 2-5*; 14-36%). 5 equiv. of peptide was found to be optimal, potentially due to the need to balance protein-bound vs solution-phase activation of the NASA species, which would in turn lead to hydrolysis. In the absence of **P1-(PyOx)$_1$** there was no labelling (*Entry 1*), demonstrating that acylation was peptide-mediated and that background reactivity of the NASA was minimal under these conditions.

No difference in conversion was observed if reactions were extended to 24 h (*Entry 6*), consistent with the limited hydrolytic stability of the intermediate oxime-acyl complex formed during catalysis, as well as the slower (but still significant) hydrolysis rate of the



Entry	pH	T / °C	Equiv. P1-(PyOx)1	Equiv. 3b	t / h	Modification / %
1	7.4	22	0	20	2	0
2	7.4	22	1.25	20	2	23
3	7.4	22	2.5	20	2	14
4	7.4	22	5	20	2	36
5	7.4	22	10	20	2	26
6	7.4	22	5	20	24	35
7	6	22	5	20	2	-
8	8	22	5	20	2	8
9	7.4	4	5	20	2	0
10	7.4	37	5	20	2	17

Figure 3: Optimisation of biotin labelling of insulin via peptide-directed catalysis with peptide **(PyOx)$_1$-P1** and NASA reagent **3b**. Modification determined by LC-MS analysis of crude reaction mixtures.

NASAs themselves. However, following these extended periods very low levels of background labeling were observed in the absence of peptide (<2%), and so 2 h was chosen as a reaction length for all future experiments.

We next looked to optimise the reaction conditions to maximise conversion without increasing the levels of non-catalysed, off-target modification. Labelling at pH 6 was not possible due to the decreased solubility of insulin at this pH, leading to a loss of protein MS signals even in the absence of peptide or NASA (*Entry 7*).[18] At pH 8, labelling efficiency decreased relative to pH 7.4, potentially due to higher levels of NASA hydrolysis, since the nucleophilicity of the target amino acid residues and PyOx catalyst would be expected to increase at higher pH (*Entry 8*). With **P1-(PyOx)$_1$,** no labeling was observed when lowering the reaction temperature to 4 °C. At 37 °C, modification was successful but less efficient, potentially due to an increased rate of NASA hydrolysis, and so room temperature (~22 °C) was used in subsequent experiments (*Entries 9-10*). However, as discussed below, we expect the optimal conditions to be peptide- and protein-dependent, due to the complex interplay between NASA activation or hydrolysis, and the subsequent labelling profiles of different amino acid residues.

Using the optimised conditions for **P1-(PyOx)$_1$**, we were also able to demonstrate labelling with NBD-NASA (**3c**). Strong fluorescent labelling of insulin was observed via in-gel fluorescence, while successful biotinylation with **3b** was also verified via anti-biotin western blotting (Supplementary information, *Fig. S5 and S7*). Notably, the sensitivity of both fluorescent imaging and western blotting allowed us to detect very low levels of background labelling at higher NASA loadings in the absence of **P1-(PyOx)$_1$** even after 2 h labelling, though this was found to be taking place at levels that were undetectable by LC-MS.

We next looked to investigate the use of catalyst **(PyOx)$_1$-P1**, sharing the same RGFFYT insulin-binding sequence but bearing the pyridinium oxime at the N-terminus. This peptide was also found to mediate successful insulin labelling, though with reduced conversion (25% conversion, see *SI Section 4*). Given the similarity in K_ds between the two peptides, we attribute this difference in labelling efficiency to the availability or reactivity of suitable amino acids in proximity to the C-terminal vs N-terminal binding site. In an attempt to improve labelling efficiency, we also synthesized a peptide bearing two pyridinium oxime catalysts at the C-terminus via a bis-functional azide, to generate **P1-(PyOx)$_2$** (***Fig. 2c***). Keijzer *et al* have previously reported that aptamers bearing two pyridinium oximes were able to enhance protein labelling relative to the monovalent analogue.[19] However, in contrast we saw a reduction in labelling efficiency (19%, see *SI Section 4*), which may be attributed to a change in the orientation or positioning of the PyOx catalyst within the bis-functional peptide.

Matrix-assisted laser desorption/ionisation MS/MS (MALDI-MS/MS) analysis was subsequently performed to identify the sites of insulin labelling for each

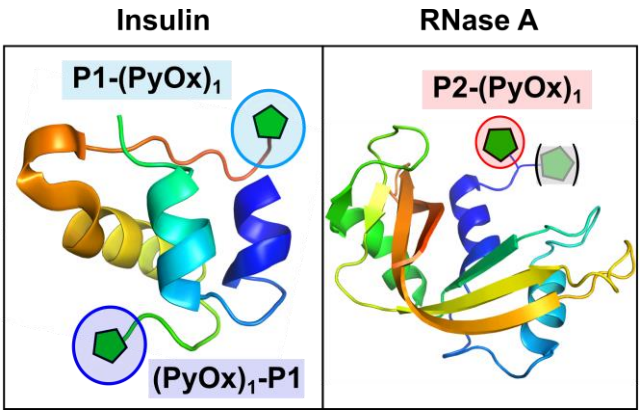


Figure 4: Sites of insulin and RNase labelling using the peptide catalysts developed in this study. For RNase A, a mix of mono- and di-labelling was observed due to modification of both the α- and ε-amines of the N-terminal lysine residue.

peptide catalyst. For N-terminal catalyst peptide **(PyOx)$_1$-P1**, labelling was found to be taking place within the first 7 residues of the N-terminal region of insulin chain B (FVNQHLC, see *SI Section 7*, ***Fig. 4***). Within this region, the N-terminal α-amine is the most likely site for labelling, though we were not able to demonstrate this unambiguously. In contrast, using **P1-(PyOx)$_1$** with the catalyst at the C-terminus of the peptide, labelling was found to take place specifically at the N-terminus of insulin chain A (see *SI Section 7,* **Fig. 4**). This validates our hypothesis that different sites on a protein can be targeted via a peptide-directed catalysis approach, by tuning the peptide structure, providing a key benefit for this approach over small molecule-mediated labelling.

A potential limitation of using peptide-based ligands for proximity catalysis is the risk of preferential self-labeling of the peptide itself, rather than the protein, if it contains nucleophilic residues. The effects of self-labelling will differ depending on the peptide, with a decrease in protein binding affinity one possibility. Alternatively, self-labelling may have a neutral effect on K_d. In this scenario, protein labelling would still be possible, albeit with a potentially reduced stoichiometry of NASA reagent. These effects are likely to be complex, with the rates of peptide-protein binding/debinding, NASA-activation, and reactivity with different amino acids on both peptide and protein contributing to the final outcome. Evidence of this was seen in control reactions in which insulin was omitted, with peptides **P1-(PyOx)$_1$** and **(PyOx)$_1$-P1**. In these controls, the peptides were incubated with hexanoic- (**3a**) or biotin-NASAs (**3b**) and analysed by LC-MS at varying time points. For **P1-(PyOx)$_1$** a new species with mass matching the acylated-peptide was observed to increase over time (see *SI Section 8*). In contrast, for **(PyOx)$_1$-P1** no modification was detected. This difference in labeling between the two peptides, sharing a common backbone sequence, suggested that at least some of the observed labeling was taking place via an intramolecular route, though the possibility of intermolecular labeling cannot be ruled out at this stage. MS/MS analysis of the labelled **P1-(PyOx)$_2$** complex allowed us to identify that the peptide was being modified within the first three residues of the peptide, RGF. Analogous results were obtained for experiments using hexanoic-NASA **3a**. The modification of arginine side-chains with NASA reagents has not previously been reported, and so it is likely that this labelling was taking place at the free α-amine of the N-terminus. Notably, in the case of **(PyOx)$_1$-P1** the pyridinium oxime was introduced at the N-terminus via a pentynoic acid linker, and the peptide did not possess an α-amine. This observation may therefore provide useful design criteria for peptide-binding catalysts in the future, particularly where a free N-terminus is not essential for protein-binding and can be e.g. acetylated.

Having validated our labelling strategy on insulin, we next looked to demonstrate translatability on an alternative protein target. The S-peptide KETAAAK-FERQHMDSSTSA, **P2**, can be proteolytically cleaved from the N-terminal domain of RNase A, wherein it retains binding capacity to the remaining parent protein, commonly referred to as RNase S, with high affinity.[20] We reasoned that a low K_d would potentially make RNase S a challenging target to work with, but hypothesized that **P2** might retain some binding affinity for RNase A, even before proteolysis. QCM-D was again used to investigate the binding of peptides tethered to the sensor via their N- or C-termini (using the thiol-labelled peptides **HS-P2** and **P2-SH** respectively). **P2-SH**, tethered via its C-terminus, was found to bind RNase A with K_d ~ 2 μM, and was therefore seen as a suitable sequence to carry forward into protein labelling experiments (Supplementary information, *Fig. S3 and S4*). In contrast, **HS-P2**, tethered via its N-terminus, showed no RNase A **binding**, indicating that functionalization with a pyridinium oxime catalyst was unlikely to be tolerated. We therefore synthesized peptide **P2-(PyOx)$_1$** and tested its ability to modify RNase A with Cy5-functionalised NASA **3d**. Pleasingly, in the presence of 5 equiv. **P2-(PyOx)$_1$** and 20 equiv. **3d** clear fluorescent labelling of RNase A was observed, with a small shift in the molecular weight of the protein observed via TSDS-PAGE (Supplementary information *Fig. S6*). NanoLC-MS/MS analysis demonstrated that modification was happening specifically at the N-terminal lysine residue, with mono- and di-modification of this amino acid observed, due to labelling at both the free α- and ε- amines (see *SI Section* 7, **Fig. *4***). This demonstrates that our peptide-directed catalysis approach can be applied to alternative protein targets, and opens up the possibility of developing a generalisable strategy for proximity-mediated protein labelling.

Conclusions

In this work we have demonstrated the use of functionalised peptide ligands as catalysts for proximity-dependent protein labelling. The use of peptide ligands provides a number of key advantages over the small molecules classically used in this context: i) labelling can be directed away from the active site of the enzyme, as epitomized by our observed N-terminal labelling of RNase A, far away from the catalytic RNA binding pocket; and ii) simple changes to the peptide-catalyst structure, enabled by the high modularity of solid-phase peptide synthesis, and demonstrated in our experiments through either N- or C-terminal functionalisation of insulin binding peptides, can alter labelling site. When coupled with the structural diversity of potential peptide ligands and the amenability of peptides to library screening against a particular protein target, this opens opportunities to tune the labelling site for a given downstream application. We envisage that our approach may therefore enable controlled, site-selective modification of otherwise challenging to label proteins, such as those of native origin that are not amenable to genetic engineering. With increasing interest in proximity-mediated catalysis in the bioconjugation community, the general approach detailed in this work can be expanded, for example to alter the amino acid labelling preference in cases where suitable nucleophilic residues are not present in proximity to the peptide binding site, enable multi-site labelling of proteins with peptides bearing two catalysts at rationally selected positions, or to promote selective modification of specific proteins in complex mixtures. We therefore believe the technology outlined in this work represents an important new addition to the bioconjugation toolbox.

ASSOCIATED CONTENT

Supporting Information

The supporting information contains all experimental details, including peptide and small molecule synthesis, protein binding and labelling, and analysis of labelled constructs.

AUTHOR INFORMATION

Author contributions

Laetitia Raynal and Joe Nabarro contributed equally to this manuscript. LR and JN performed and analysed protein labelling experiments. LR, JN, and SLM synthesized the required peptides. LR synthesized all small molecule reagents. LMM performed QCM-D of the synthesised peptides. AAD performed MALDI-MS and nanoLC-MS on modified proteins. SDJ, MAF, and CDS supervised the study. CDS developed and managed the study, and wrote the manuscript. All authors contributed to the editing of the manuscript.

Competing interests

The authors declare no competing financial interests.

ACKNOWLEDGMENT

Anaïs Sanchez is thanked for the synthesis of precursors. LR and CDS are grateful to the Rosetrees Trust for PhD Studentship funding (agreement A2413). JN and MAF acknowledge the BBSRC for PhD studentship support through the White Rose DTP (2272649). LMM and SDJ thank the University of York for funding to purchase the QCM-D used, and a Wellcome Trust Technology Development Grant (221349/Z/20/Z). CDS and SLM acknowledge generous support through a Wellcome Trust Career Development Award (225257/Z/22/Z). We thank Dr Ed Bergstrom and The York Centre of Excellence in Mass Spectrometry for support with peptide mass spectrometry. This centre was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from the EPSRC (EP/K039660/1; EP/M028127/1).

REFERENCES

1 C. D. Spicer and B. G. Davis, *Nat. Commun.*, 2014, **5**, 4740.
2 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis, *Chem. Sci.*, 2015, **6**, 50–69.
3 D. Hymel and F. Liu, *Asian J. Org. Chem.*, 2021, **10**, 38–49.
4 K. Shiraiwa, R. Cheng, H. Nonaka, T. Tamura and I. Hamachi, *Cell Chem. Biol.*, 2020, **27**, 970–985.
5 M. R. Mortensen, M. B. Skovsgaard and K. V. Gothelf, *ChemBioChem*, 2019, **20**, 2711–2728.
6 H. A. Beard, J. R. Hauser, M. Walko, R. M. George, A. J. Wilson and R. S. Bon, *Commun. Chem.*, 2019, **2**, 133.
7 Y. Wang, R. Zhao, C. Wan, X. Guo, F. Yang, Z. Hou, R. Wang, S. Li, T. Feng, F. Yin and Z. Li, *Org. Lett.*, 2022, **24**, 7205–7209.
8 Y. Zhang, Y. Liang, F. Huang, Y. Zhang, X. Li and J. Xia, *Biochemistry*, 2019, **58**, 1010–1018.
9 M. Kaminska, P. Bruyat, C. Malgorn, M. Doladilhe, E. Cassar-Lajeunesse, C. Fruchart Gaillard, M. De Souza, F. Beau, R. Thai, I. Correia, A. Galat, D. Georgiadis, O. Lequin, V. Dive, S. Bregant and Devel, Laurent, *Angew. Chem. Int. Ed.*, 2021, **60**, 2–10.
10 W. Jaroszewicz, J. Morcinek-Orłowska, K. Pierzynowska, L. Gaffke and G. Węgrzyn, *FEMS Microbiol. Rev.*, 2022, **46**, fuab052.
11 Y. Koshi, E. Nakata, M. Miyagawa, S. Tsukiji, T. Ogawa and I. Hamachi, *J. Am. Chem. Soc.*, 2008, **130**, 245–251.
12 J. Ohata and Z. T. Ball, *J. Am. Chem. Soc.*, 2017, **139**, 12617–12622.

13 T. Hayashi and I. Hamachi, *Acc. Chem. Res.*, 2012, **45**, 1460–1469.
14 T. Tamura, Z. Song, K. Amaike, S. Lee, S. Yin, S. Kiyonaka and I. Hamachi, *J. Am. Chem. Soc.*, 2017, **139**, 14181–14191.
15 V. P. Knutson, *J. Biol. Chem.*, 1988, **263**, 14146–14151.
16 H. Q. Yu, X. Y. Dong and Y. Sun, *Chromatographia*, 2004, **60**, 379–383.
17 H. L. Chiang, S. T. Ngo, C. J. Chen, C. K. Hu and M. S. Li, *PLoS ONE*, 2013, **8**, e65358.
18 F. J. Link and J. Y. Y. Heng, *Cryst. Growth Des.*, 2022, **22**, 3024–3033.
19 J. F. Keijzer, J. Firet and B. Albada, *Chem. Commun.*, 2021, **57**, 12960–12963.
20 J.-S. Kim and R. T. Raines, *Protein Sci.*, 1993, **2**, 348–356.