Selective N-terminal modification of peptides and proteins using acyl phosphates

Laura Rodríguez Pérez[1], Thomas A. King[1], William Finnigan[1], Antonio Angelastro[1], William R. F. Goundry[2], and Sabine L. Flitsch[1]

[1] Department of Chemistry, The University of Manchester, Manchester Institute of Biotechnology, 131 Princess Street, Manchester, M1 7DN, UK.

[2] Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Macclesfield, UK.

* Corresponding author: Prof. Sabine L. Flitsch, sabine.flitsch@manchester.ac.uk

Selective acylation of the N-terminus over side-chains in peptides and proteins is a highly desirable but challenging reaction in chemical biology. Here we report a biomimetic approach using enzymatic *in situ* activation of carboxylic acids with ATP to generate reactive acyl-adenosine phosphates, which display high selectivity for the N-termini of peptides and proteins, including pharmaceutically relevant liraglutide, insulin and glucagon. The acylation tolerates a range of unsubstituted and substituted fatty acids including di-acids, thus making it suitable for N-terminal biorthogonal labelling strategies.

Over the past decade, the selective modification of proteins and peptides has taken centre stage in the field of chemical biology [1–3]. The ever-expanding toolbox of bioconjugation methods has proven useful for a myriad of applications[4,5], from the development of new targeted delivery strategies, such as antibody-drug conjugates for cancer treatments[6,7], to peptides and insulin derivatives for the treatment of diabetes[8,9]. Bioorthogonal modification of peptides is also important in proteomics studies as an essential facet of labelling strategies[10]. Among all the functional groups in proteins, the N-terminus has privileged properties over other amino acid residues, partly due to a significantly distinct pK_a value[11]. This difference in reactivity has been recognised by several groups and used for selective N-terminal conjugation using 2-pyridinecarboxyaldehydes[12], 2-ethynylbenzaldehydes[13], or phenyl esters[14]. Alternatively, there are many examples of N-terminal peptide tags introduced through protein engineering, which can be selectively modified by enzymes[15]. However, despite many efforts, bioconjugation methods that can generate stable linkages on native N-termini directly with free acids remain elusive[16]. Here, we address this challenge by reporting the *in situ* enzymatic reagent activation (ERA) of a range of natural and functionalised fatty acids to acyl-adenosyl monophosphates (acyl-AMP), showing their selectivity towards the N-terminus of peptides and proteins.

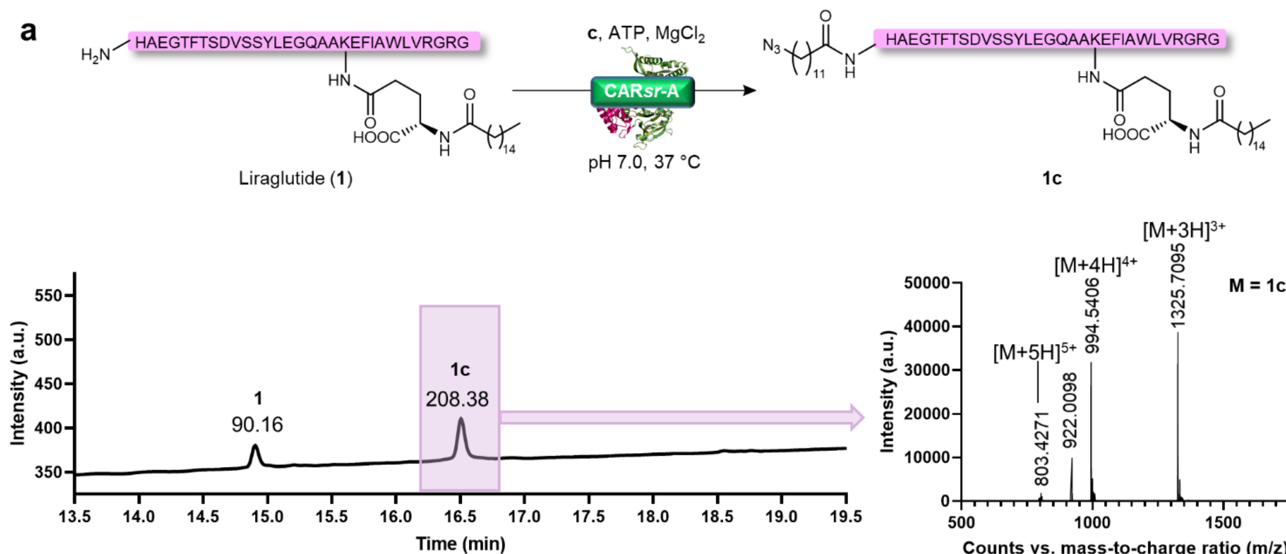
Acyl-adenosyl monophosphates are common intermediates in many metabolic pathways, but their chemistry has not been exploited because their synthesis is challenging and enzymatic activation limited to their natural substrates. Our laboratory has previously shown that the adenylation domain (-A) of carboxylic acid reductase (CAR) from *Segniliparus rugosus* (CAR*sr*-A) can be used to catalyse the formation of acyl-SCoA of a wide variety of acids using ATP[17]. It was envisaged that, in the absence of CoASH, the acyl-AMP intermediate could be used for the selective modification of amines in proteins. Its preference for the N-terminus was anticipated due to its privileged nature over the other less nucleophilic functional groups.

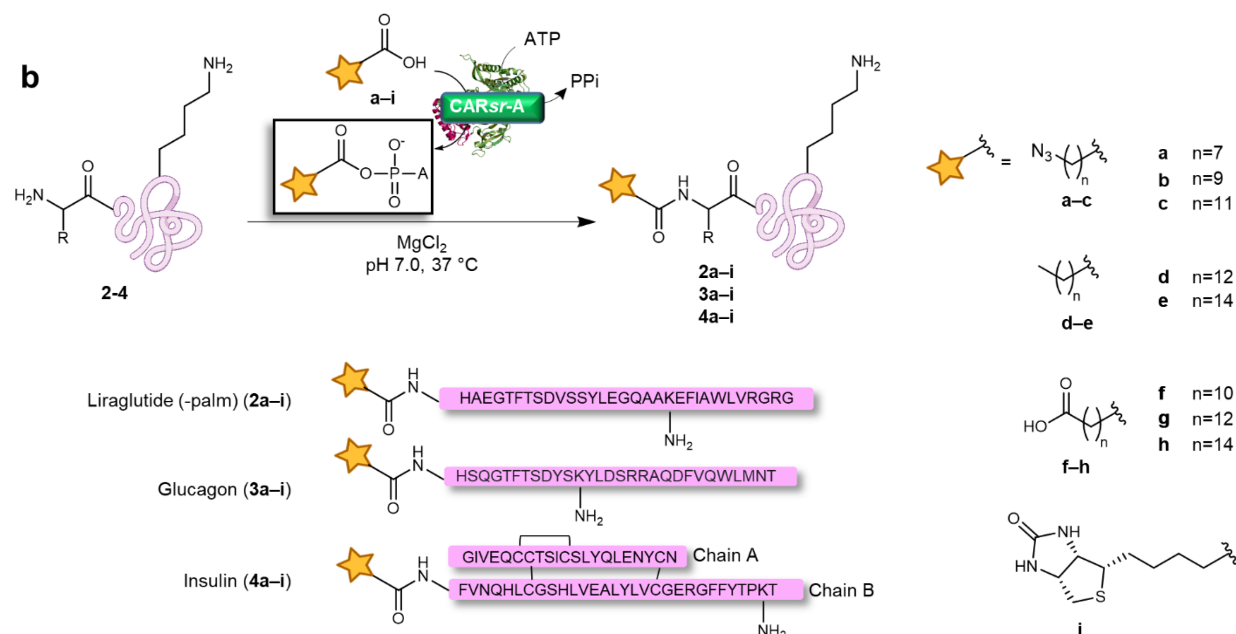
In the first instance, the peptide liraglutide (**1**, **Figure 1a**) was used as a substrate containing only one amino group on the N-terminus. This peptide and respective acylated derivatives have gained increased attention due to their use in the treatment of type 2 diabetes and obesity[8]. Treatment of **1** with acid **c** and the ERA reagents (ATP, CAR*sr*-A) led to formation of the N-acylated product (**1c**), with 70% conversion as measured by LC-MS with no other product detected.

Following on from these promising results, the selectivity of the ERA method was tested with peptides containing both N-terminal and lysyl amino groups. The influence of reaction conditions on N-terminal selectivity and conversion was systematically investigated by performing Definitive Screening Design (DSD)[18] on a liraglutide analogue with a free lysine (**2**) and acid **b** (**Supplementary Information Figure 2**). The DSD model for selectivity showed that the pH of the reaction mixture, as well as concentrations of CAR*sr*-A and peptide **2** were the most significant variables. Lower concentrations of CAR*sr*-A, a pH value of 7.0 and peptide concentration at 300 µM yielded the best selectivity for the N-terminus. The selectivity of the reaction for N-terminal over lysine side-chain acylation was determined by LC-MS/MS analysis of product **2b** (**Supplementary Figure 34**). A second model, predicting conversion, suggested that the most significant factors were the concentration of peptide, CAR*sr*-A and acid, with optimal values of 183 µM, 10 µM, and 8 mM respectively. Overall, validation experiments indicated that pH 7.0 also favoured conversion (**Supplementary Figure 2**). To further optimise the conversions to the product, a response surface was generated considering CAR*sr*-A and acid concentrations (**Supplementary Figure 3**). After the models for conversion and selectivity were developed, it was determined that the best conditions for this reaction were 50 mM HEPES buffer pH 7.0, 190 µM **2**, 13 mM **b**, 10 mM $MgCl_2$, 8 µM CAR*sr*-A, and 5 mM ATP at 37 °C.

To explore the applicability of the ERA strategy to other peptides and proteins, the acylation of glucagon (**3**) and insulin (**4**) was investigated. Acylation of glucagon with acid **b** was shown to primarily target the N-terminus **(Supplementary Figures 25 and 35).** Selective modification of insulin would be of particular interest due to the use of insulin derivatives as biopharmaceuticals[19]. Insulin contains three potential modification sites, two N-terminal α-amines and a lysine ε-amine. Interestingly, after ERA treatment of insulin, LC-MS/MS analysis of the product showed that the acylation with **b** was directed to the N-terminus of chain B (**Supplementary Figures 7 and 36**). This is expected based on the relative pKa values of the chain B N-terminus (7.1) and those of the chain A N-terminus (8.4) and the lysine (11.1)[20].

https://doi.org/10.26434/chemrxiv-2023-jm6sb **ORCID:** https://orcid.org/0000-0002-3382-4407 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0





% Selectivity (% Conversion)

	a	b	c	d	e	f	g	h	i
2	90 (27)	94 (94)	82 (67)	>99 (26)	>99 (20)	>99 (39)	>99 (44)	>99 (33)	>99 (12)
3	>99 (10)	84 (56)	78 (71)	>99 (5)	>99 (8)	88 (25)	85 (15)	>99 (16)	>99 (5)
4	92 (34)	87 (87)	89 (89)	>99 (44)	>99 (3)	77 (45)	75 (63)	79 (63)	75 (12)

Figure 1. Enzymatic reagent activation (ERA) strategy for the N-terminal acylation of peptides and proteins *via* acyl phosphates. a, N-terminal acylation of peptide **1** with acid **c** using the ERA strategy as confirmed by LC-MS. Reaction conditions: 50 mM HEPES buffer pH 7.0, 190 µM **1**, 13 mM **c**, 10 mM $MgCl_2$, 8 µM CAR*sr*-A, 5 mM ATP, 37 °C, 24 h. Areas are indicated above each peak. **b,** Selective acylation of the N-termini of peptides **2-4** with acids **a–i**. % Selectivity refers to the percentage of N-terminal peptides **2a–i, 3a–i,** and **4a–i** over all acylated peptide products. Overall conversions of N-terminal acylation are shown in brackets.

https://doi.org/10.26434/chemrxiv-2023-jm6sb ORCID: https://orcid.org/0000-0002-3382-4407 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

Further DSD experiments carried out on substrates **3** and **4** with acids **b** and **h** (**Supplementary Figures 4–6**) showed that in all cases a pH value of 7.0 was optimal. Furthermore, the same effects of acid substrate and CAR*sr* concentrations were observed for the selectivity model as for the previous DSD on peptide **2**. In general, optimal reaction conditions were 50 mM HEPES buffer pH 7.0, 190–300 μM peptide **2–4**, 13 mM acid **a–i**, 10 mM $MgCl_2$, 8–35 μM CAR*sr*-A, 5 mM ATP (**Supplementary Methods**).

With this ERA reaction protocol in hand, the substrate scope was compared for different acids (**a-i**) and peptides (**2–4**) (**Figure 1b**) both in terms of selectivity and conversion. In addition to fatty acids **d** and **e,** azido acids **a–c** enabled the introduction of 'clickable' groups into peptides and proteins for further chemical bioconjugation[1]. Dicarboxylic acids **f–h** were substrates of interest given the additional demand on selectivity without protecting groups. And finally, the biotin derivative **i** was chosen to investigate direct introduction of affinity labels. The selectivity of the reaction was investigated by LC-MS and in all cases the main product was found to be the result of N-terminal acylation with additional lysine acylation observed for some substrates (**Supplementary Figures 7–33**).

A summary of all the results of the ERA-mediated acylation reactions are presented in **Figure 1b**. In general, good to high selectivity was observed with many peptide and acid substrates and functionalisation was achieved in all cases. Azido fatty acids **b** and **c** yielded medium to high conversions across all substrates (56–94%)**.** The yield of biotinylation was modest (5–12%), suggesting that larger labels are more challenging for the reaction. Selectivity for the N-terminal amino group was good to excellent. In many cases, N-terminal acylation was the only product identified by LC-MS (>99% selectivity). The results for acids **f–h** led to the respective monoacylated products, showing that the reaction is successful without the need for protecting the second carboxyl group. Interestingly, no macrocyclisation was observed.

To our knowledge, this is the first time that a generic library of acyl-adenosyl monophosphates has been enzymatically generated *in situ* for the modification of peptides and proteins. The library was assessed in terms of their chemical reactivity and selectivity, and their preference for the N-terminal modification was discovered. The additional selectivity we observe for the B chain over the A chain of insulin suggests that lower pK_a values of the N-terminal amine, such as those of the N-terminus of His and Phe, favour the reaction. Therefore it is predicted that other N-terminal amino acids that lower the α-amine pK_a such as Lys, Asn, Arg, Tyr or Met[11,21] will also yield good conversions. Altogether, the ERA technology described here represents an *in vitro* biomimetic method to produce peptide and protein conjugates with high selectivity for N-termini over lysine side chains. ERA is complementary to other bioorthogonal methods in that free acids can be used directly, with no need for prior chemical activation. Moreover, these reagents have low toxicity and are easily accessible, and no recognition sequence is required in the peptide substrate. As such, the ERA method should find wider applications as a universal bioorthogonal tool for labelling of peptides and proteins and for the synthesis of stable protein conjugates in biotechnology.

Acknowledgements

This study was funded by the EPSRC, BBSRC, and AstraZeneca plc under the Prosperity Partnership EP/S005226/1. The authors thank Reynard Spiess for support with MS analyses and Christian Schnepel for making the CAR*sr*-A plasmid. Figure was partially created using Biorender.com.

References

1. Sletten, E. M., Bertozzi, C. R., Bertozzi, C. R. & Sletten, E. M. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009).
2. Krall, N., Da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113 (2016).
3. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114**, 4764–4806 (2014).
4. Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L. & Bernardes, G. J. L. Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **3**, 147–171 (2019).
5. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**, 1–14 (2014).
6. Dumontet, C., Reichert, J. M., Senter, P. D., Lambert, J. M. & Beck, A. Antibody–drug conjugates come of age in oncology. *Nat. Rev. Drug Discov.* (2023).
7. Drago, J. Z. Unlocking the potential of antibody–drug conjugates for cancer therapy. *Nat. Rev. Clin. Oncol.* **18**, 327–344 (2021).
8. Lau, J. *et al.* Discovery of the once-weekly glucagon-like peptide-1 (GLP-1) analogue semaglutide. *J. Med. Chem.* **58**, (2015).
9. Coskun, T. *et al.* LY3298176, a novel dual GIP and GLP-1 receptor agonist for the treatment of type 2 diabetes mellitus: From discovery to clinical proof of concept. *Mol. Metab.* **18**, 3–14 (2018).
10. Parker, C. G. & Pratt, M. R. Click chemistry in proteomic investigations. *Cell* **180**, 605–632 (2020).
11. Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **13**, 697–705 (2017).
12. Macdonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11**, 326–331 (2015).
13. Deng, J. *et al.* N-Terminal selective modification of peptides and proteins using 2-ethynylbenzaldehydes. *Commun. Chem.* **3**, 1–9 (2020).

https://doi.org/10.26434/chemrxiv-2023-jm6sb **ORCID:** https://orcid.org/0000-0002-3382-4407 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

14. Mikkelsen, J. H., Gustafsson, M. B. F., Skrydstrup, T. & Jensen, K. B. Selective N-terminal acylation of peptides and proteins with tunable phenol esters. *Bioconjug. Chem.* **33**, 625–633 (2022).
15. Heal, W. P., Wright, M. H., Thinon, E. & Tate, E. W. Multifunctional protein labeling via enzymatic N-terminal tagging and elaboration by click chemistry. *Nat. Protoc.* **7**, 105–117 (2012).
16. Tantipanjaporn, A. & Wong, M. K. Development and Recent Advances in Lysine and N-Terminal Bioconjugation for Peptides and Proteins. *Molecules* **28**, (2023).
17. Schnepel, C. *et al.* Thioester-mediated biocatalytic amide bond synthesis with in situ thiol recycling. *Nat. Catal.* **6**, 89–99 (2023).
18. Lubberink, M. *et al.* One-step biocatalytic synthesis of sustainable surfactants by selective amide bond formation. *Angew. Chem. Int. Ed.* **61**, e202205054 (2022).
19. Kramer, C. K., Retnakaran, R. & Zinman, B. Insulin and insulin analogs as antidiabetic therapy : A perspective from clinical trials. *Cell Metab.* **33**, 740–747 (2021).
20. Fryszkowska, A. *et al.* A chemoenzymatic strategy for site-selective functionalization of native peptides and proteins. *Science* **376**, 1321–1327 (2022).
21. Sereda, J., Mant, C. T., Quinn, A. M. & Hodges, S. Effect of the a-amino group on peptide retention behaviour in reversed-phase chromatography Determination of the pK , values of the a-amino of 19 different N-terminal amino acid residues. *J. Chromatogr.* **646**, 17–30 (1993).

https://doi.org/10.26434/chemrxiv-2023-jm6sb **ORCID:** https://orcid.org/0000-0002-3382-4407 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0