Chemistry–A European Journal

Research Article
doi.org/10.1002/chem.202202377

Chemistry Europe
European Chemical
Societies Publishing

www.chemeurj.org

Interactive Bioconjugation at *N*-Terminal Cysteines by Using *O*-Salicylaldehyde Esters towards Dual Site-Selective Functionalization

Maria J. S. A. Silva,*[a] Rafaela A. N. Cavadas,[a] Hélio Faustino,[a, b] Luís F. Veiros,[c] and Pedro M. P. Gois*[a]

Dedicated to Prof. Carlos A. M. Afonso on the occasion of his 60th birthday

Abstract: *N*-terminal Cys modification has been intensively studied to produce homogeneous bioconjugates essentially through two modes of reaction: reversible modification with the equilibrium shifted towards the formation of the desired conjugate or stable and irreversible conjugates. Herein, we report a new method of *N*-terminal cysteine modification using *O*-salicylaldehyde esters (OSAEs) through fast conjugation and irreversible deconjugation. These reagents can rapidly react with *N*-terminal Cys at low-micromolar concentration to form thiazolidines with subsequent hydrolysis of the ester moiety to the phenolic derivative. These phenolic thiazolidines can be hydrolyzed at acidic pH (≈ 4.5) to recover the intact *N*-terminal Cys. Bioconjugation reactions using OSAEs offer controlled reversibility to as act as a protecting group for *N*-terminal cysteines, allowing the modification of in-chain residues without perturbing the *N*-terminal Cys, which can then be deprotected and used as a conjugation site.

Introduction

The modification of a biomolecule (e.g., a peptide, protein, hormone, nucleic acids) with a functional cargo enables the construction of bioconjugates with the combined properties of both components. In recent years, these technologies have become a central strategy for the synthesis of hybrid materials applicable in many different areas.[1]

[a] Dr. M. J. S. A. Silva, R. A. N. Cavadas, Dr. H. Faustino, Dr. P. M. P. Gois
Research Institute for Medicines (iMed.ULisboa)
Faculty of Pharmacy, Universidade de Lisboa
Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisboa (Portugal)
E-mail: maria.silva7@campus.ul.pt
          pedrogois@ff.ulisboa.pt

[b] Dr. H. Faustino
Centre Bio R&D Unit
Association BLC3 – Innovation and Technology Campus
3405-155 Oliveira do Hospital (Portugal)

[c] Dr. L. F. Veiros
Centro de Química Estrutural, Institute of Molecular Sciences
Departamento de Engenharia Química, Instituto Superior Técnico
Universidade de Lisboa
Av. Rovisco Pais, 1049-001 Lisboa (Portugal)

Bioconjugation methods evolved to offer control over the site of the functionalization enabling the design of constructs with improved homogeneity. The low abundance of cysteines combined with the nucleophilicity of the sulfhydryl group, made this residue the preferred target for the development of site-selective technologies.[2] However, many of the reactions targeting the reactivity of the sulfhydryl group proved unable to differentiate between different reduced Cys residues present on the peptide chain.[3] This poor selectivity has limited the access to well-defined multifunctionalized bioconjugates. Therefore, many efforts have been developed to establish orthogonal methods for Cys functionalization. In this context, *N*-terminal cysteines emerged as a popular target for site-specific bioconjugation, using reagents that target selectively the 1,2-aminothiol group.[4] These reagents typically offer high chemoselectivity and rapid kinetics, though they offer poor control over the reversibility of the installed handle. For these reasons, these technologies have limited applicability in the construction of complex bioconjugates that require a sequence of interactive modifications (Figure 1). Hence, the discovery of new reagents that provide site-selective bioconjugation reaction with fast kinetics and simple mechanisms to control the deconjugation process are much needed to design the next generations of multifunctionalized bioconjugates.

Aldehydes are one of the most widely explored handles for the selective conjugation with *N*-terminal cysteines,[5] (Figure 1) and several studies have been conducted to improve the kinetics of the thiazolidine formation,[6] that require the use of a large excess of the aldehyde or long reaction times. Aware of these limitations, Gois and Gao independently reported the use of 2-formyl phenyl boronic acids to modify *N*-terminal Cys with

Chem. Eur. J. **2022**, *28*, e202202377 (1 of 6)

© 2022 The Authors. Chemistry - A European Journal published by Wiley-VCH GmbH

Chemistry–A European Journal

Research Article
doi.org/10.1002/chem.202202377

Chemistry Europe
European Chemical
Societies Publishing
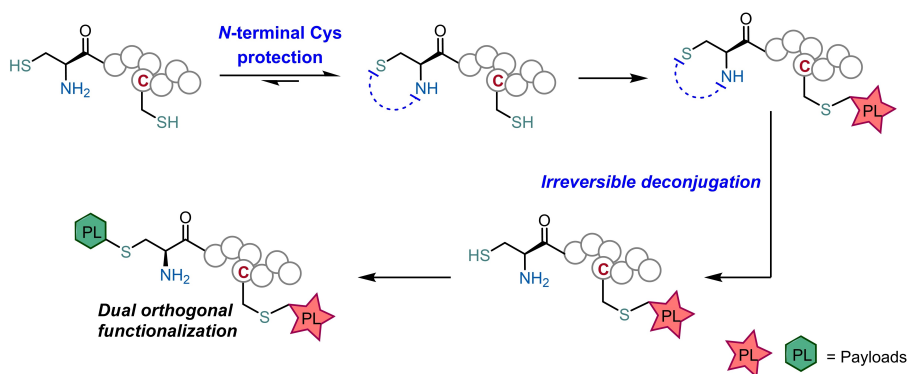
Interactive site-selective modifications

Figure 1. The aim of this work is to explore a selective protecting group for *N*-terminal Cys that irreversibly releases the starting peptide on demand upon treatment with a mild and innocuous trigger, while allowing the interactive modification of different Cys.

unprecedented fast reaction rates.[7] Furthermore, the stability of these new boronated thiazolidines was shown to be tunable. The addition of an α-nucleophile[7a,8] or lowering the pH of the reaction mixture to 2[7b] rapidly promotes thiazolidine hydrolysis. Moreover, Gao and Lee showed that *N*-acyl transfer generates stable *N*-acyl phenolic thiazolidines.[9] Despite these very attractive properties, boronated and *N*-acylated thiazolidines are either too reversible or too stable, which are difficult to use as a *N*-terminal protection group. Based on this, we envisioned that by preventing the *N*-acyl transfer would generate phenolic thiazolidines with tunable stability properties to be used in protection/deprotection sequences.

Considering these precedents, we set out to develop a new method for *N*-terminal Cys modification by using a protection-deprotection strategy that enables the orthogonal differentiation of *N*-terminal versus in-chain cysteine residues. Here, *O*-salicylaldehyde esters (OSAE) were explored as efficient reagents that rapidly form thiazolidines in mild, aqueous and dilute conditions, while being irreversibly deconjugated under acidic conditions without requiring external triggers like metal catalysts, oxidants or nucleophiles.

Results and Discussion

To test this idea, 2-formylphenyl acetate **1** was prepared and used in the bioconjugation with Laminin fragment (10 μM) as a model peptide, featuring a *N*-terminal Cys residue (Scheme 1). Under these conditions (20 mM ammonium acetate at pH 7.0), the reaction generated a phenolic thiazolidine very efficiently in 1 h with little excess of reagent **1**. To our surprise, no acetylation of the thiazolidine was observed with this reagent, while the conjugate remained mostly intact in solution over 24 h, indicating that despite having no acetyl group this phenolic thiazolidine is fairly stable (see the Supporting Information). Interestingly, in the same reaction conditions, benzaldehyde **2**, salicylaldehyde **3**, methyl 2-formylbenzoate **4**, 2-acetylphenyl acetate **5** failed to produce the thiazolidine heterocycle (≤ 15 % conversion; Figures S1–S8 in the Supporting
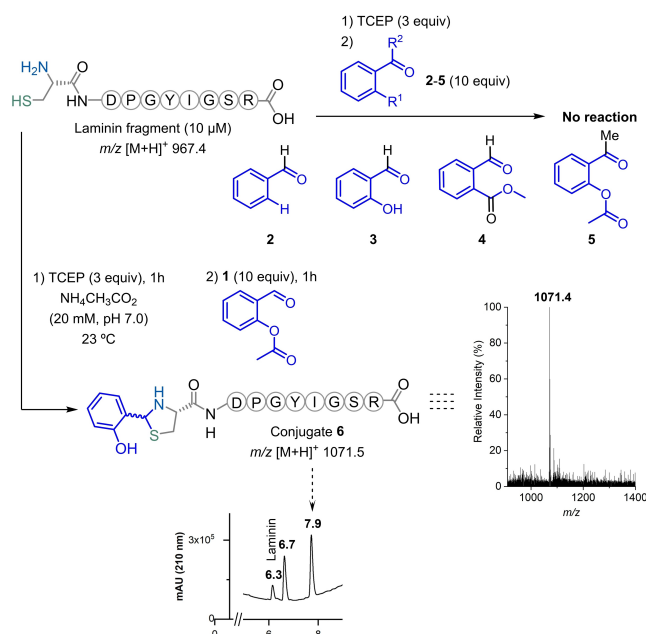
Scheme 1. Laminin bioconjugation screening with different aromatic aldehydes that only showed reaction with 2-formylphenyl acetate **1**, under aqueous conditions. The reactions' progress was monitored by ESI[+]-MS direct injection and LC-HRMS. The peak at 6.7 min is the TCEP adduct with 2-formylphenyl acetate **1**.

Information). Together, these results suggest that the acetyl function in **1**, is pivotal in the reaction mechanism and contributes to accelerate the condensation.

Intrigued by the role of the ester moiety in this reaction, we studied the mechanism of the reaction. According to DFT calculations,[10,11] the reaction is initiated by the attack of the *N*-terminal amine to the aldehyde providing the hemiaminal. Then, a transesterification process leads to the phenol OH group followed by acetic acid elimination and iminium formation. Finally, the attack of the thiol to the iminium C-atom forms the final thiazolidine product **7**. The overall process is

Chemistry–A European Journal

Research Article
doi.org/10.1002/chem.202202377

Chemistry Europe
European Chemical
Societies Publishing

exergonic ($\Delta G = -17$ kcal mol^{-1}) and has a barrier of $\Delta G^{\ddagger} = 21$ kcal mol^{-1} (Scheme 2).

More indications that support the proposed mechanism were obtained by ^{1}H NMR analysis of the reaction of 2-formylphenyl acetate with ethanethiol **8**, butylamine **9**, and Cys **10** (Scheme 3). In this set of experiments, we observed that the alkylamine **9** quickly promoted imine formation, followed by complete hydrolysis within 2 h to the starting materials. Also, 2-formylphenyl acetate **1** failed to react with the ethanethiol while, in the presence of Cys, rapidly formed thiazolidine **7** as a mixture of diastereoisomers.

According to ^{1}H NMR analysis, the half-life of hydrolysis of *O*-ester **1** in deuterated KP$_i$ 50 mM, pH 7/[D$_6$]DMSO (10:1) is
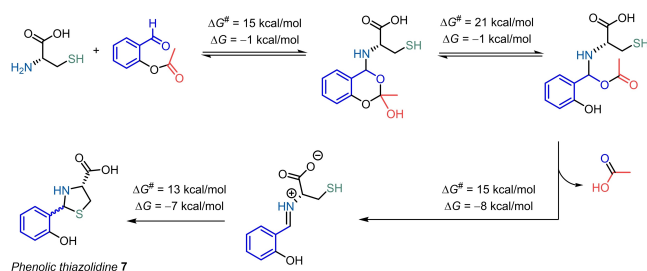
Scheme 2. OSAE reaction mechanism with N-terminal Cys to form thiazolidine **7**. Free energy values were calculated by DFT.
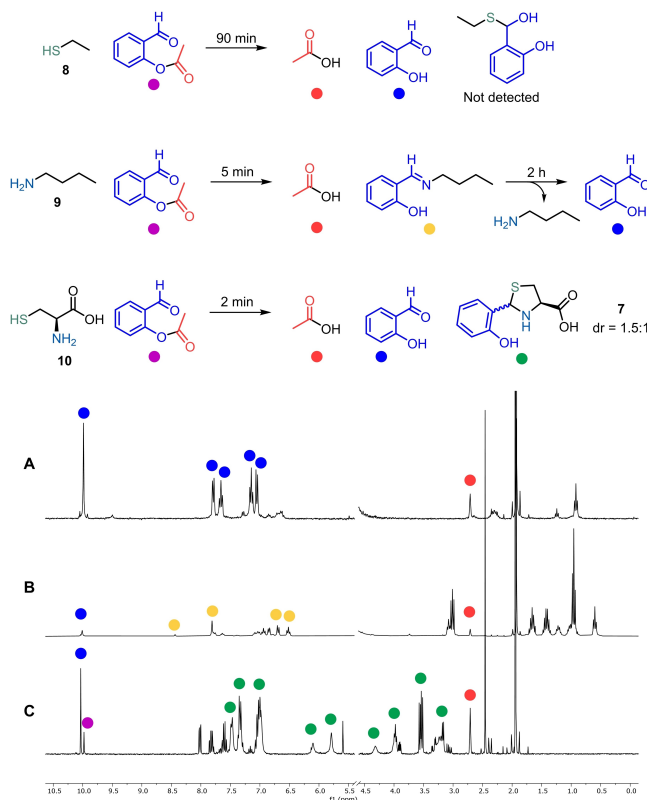

Scheme 3. Mechanism elucidation by ^{1}H NMR analysis. General conditions: Nucleophile (1 equiv) and 2-formylphenyl acetate **1** (1.2 equiv) in KP$_i$ 50 mM D$_2$O/[D$_6$]DMSO 10:1, pD 7.0. A) ethanethiol **8**, 90 min; B) butylamine **9**, 5 min; C) Cys **10**, 2 min.
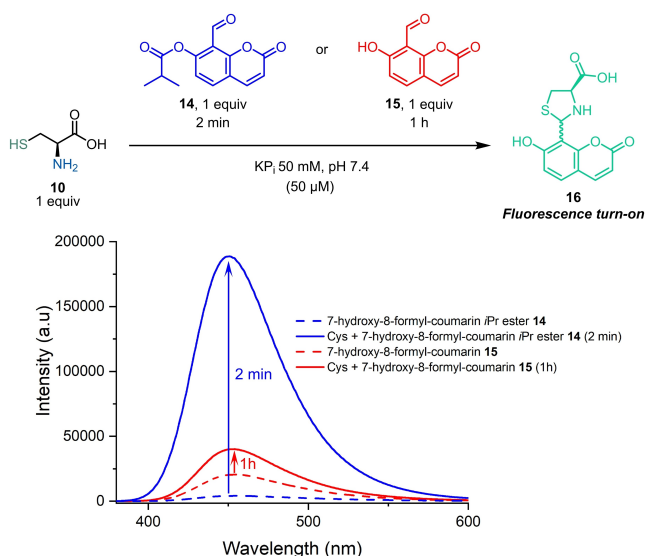
15 min which can compete with the thiazolidine formation. Therefore, different esters were prepared and tested under the same conditions previously used for Laminin fragment bioconjugation. As show in Scheme 4, we observed an increased stability in the series: Me < *i*Pr < *t*Bu < Ph, though the most stable compounds also displayed a poor solubility in the aqueous mixture (Scheme 4A). Nevertheless, **1**, **11** and **12** OSAEs efficiently conjugated with Laminin to provide the phenolic thiazolidine conjugate **6** (Scheme 4B). Only the reaction of Laminin with 2-formylphenyl pivalate **13** showed low conversion and the presence of the *N*-acetylated Laminin product,[9b] suggesting that bulkier esters undergo a different reaction mechanism, likely due to steric hindrance (see the Supporting Information). From this study, 2-formylphenyl isobutyrate **11** was chosen due to a compromise between stability and solubility.

To further corroborate the influence of the acetyl moiety in the OSAEs reactivity, we prepared a 7-hydroxy-8-formyl-coumarin that gains fluorescence upon reaction with Cys.[12] Therefore, coumarin *i*Pr ester **14** and the phenolic derivative **15** were synthesized and tested in the reaction with equimolar amount Cys. As shown in Scheme 5, only the ester **14** displayed a significant fluorescence increase in the presence of Cys (47-fold increment in fluorescence emission after only 2 min), confirming the influence of the ester in the thiazolidine formation.

Once we had established the rapid formation of the phenolic thiazolidine, we studied the stability of this heterocycle in buffers at different pH levels. To avoid high concentrations and to determine the stability at the micromolar range, extracted ion chromatograms (EIC) from LC-MS were used. While at pH 9 and 7 the hydrolysis half-time is of 3.3 h and 9 h
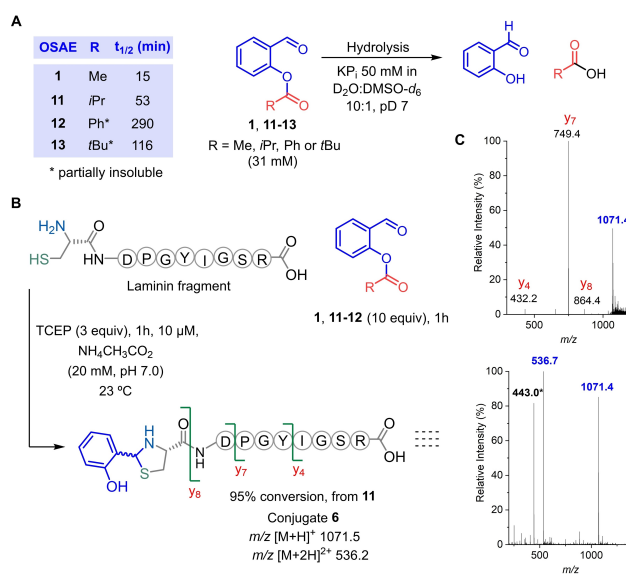
Scheme 4. Influence of different *O*-ester groups A) on the hydrolytic stability of OSAEs and B) on the bioconjugation performance with Laminin peptide. Conversion calculated for 2-formylphenyl isobutyrate **11** reaction with Laminin fragment based on EIC of the most intense peak for the starting peptide. C) ESI$^+$-MS analysis to the reaction after 1 h: *peak *m/z* 443.0 is a TCEP adduct with 2-formylphenyl isobutyrate **11** (bottom); MS fragmentation of *m/z* 1071.4 peak (top).
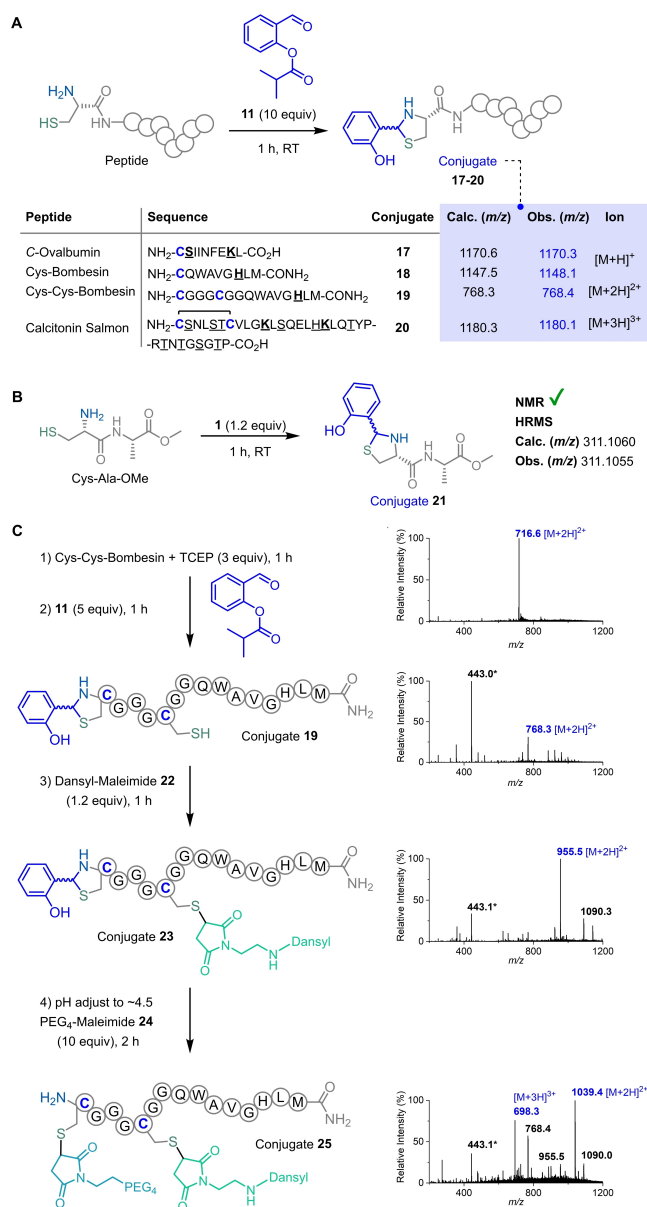
Chemistry–A European Journal

Research Article
doi.org/10.1002/chem.202202377

Chemistry Europe
European Chemical
Societies Publishing

Scheme 5. Fluorometric assay of reactions of OSAE coumarin derivative **14** vs. 7-hydroxy-8-formyl-coumarin **15** with equimolar amounts of Cys showing the enhanced rate towards the formation of thiazolidine fluorescent product **16**.

Scheme 6. A) Peptide scope to further validate the chemospecificity towards *N*-terminal Cys. General conditions: peptide (10 μM) in ammonium acetate (20 mM, pH 7.0), TCEP (3 equiv), 1 h, followed by addition of 10 equiv. of 2-formylphenyl isobutyrate **11**. After 1 h, the reaction mixture was analyzed by ESI$^+$-MS. B) Cys-Ala-OMe dipeptide (20 mM) in KPi 50 mM in D$_2$O, pD 7, 2-formylphenyl acetate **1** (1.2 equiv) in [D$_6$]DMSO. The reaction was monitored by NMR, and the diastereoisomer mixture of conjugate **21** was characterized by NMR and ESI$^+$-HRMS. C) Dual-orthogonal modification of Cys-Cys-Bombesin peptide by *N*-terminal Cys protection-deprotection with OSAE strategy for the precise installation of two different functionalized maleimides (with a dansyl fluorophore **22** and a PEG chain **24**). Conditions: Cys-Cys-Bombesin (233 μM) in ammonium acetate (20 mM, pH 7.0) at 23 °C. *peak *m/z* 443.1 is a TCEP adduct with 2-formylphenyl isobutyrate **11**.

respectively, at mild acidic pH (4.5) the hydrolysis occurred at a higher rate, with a half-life of only 42 min (Scheme 4A). These results indicate that the conjugation can be performed at neutral pH and, if needed, rapidly reverted at pH 4.5, providing an easy pH-dependent platform for *N*-terminal Cys protection without requiring external reagents like oxidants, methoxyamine or metal catalysts. In addition to this, the released salicylaldehyde is incapable to revert the mechanism back to the thiazolidine as previously demonstrated under these conditions (Schemes 1 and 5), offering superior control than boronated thiazolidines.[5,13]

Based on these results, a library of peptides was used to study the bioconjugation scope and selectivity towards *N*-terminal Cys (Scheme 6A). Apart from the Laminin fragment, Cys-Bombesin and *C*-Ovalbumin peptide containing a lysine residue that potentially reacts with the OSAEs, also cleanly afforded the phenolic thiazolidines **17** and **18** in conversions over 95% in 1 h. Cys-Cys-Bombesin, presenting a *N*-terminal and an in-chain cysteines also proceeded to form a single new product exclusively modified at the *N*-terminal position (91% conversion). The only exception was Calcitonin Salmon, where the standard protocol only achieved 56% conversion, possibly due to the easy oxidation of the cysteines to form an intramolecular disulfide bridge (see the Supporting Information). On the other hand, F3 peptide that contains a *C*-terminal cysteine and eight Lys residues, one of them at the *N*-terminal position failed to be modified with OSAE **11**, further confirming the selectivity for the *N*-terminal Cys.

In order to further elucidate the structure of the yielded phenolic thiazolidines, Cys-Ala-OMe dipeptide was used to monitor the reaction progress with 2-formylphenyl acetate **1** through ^{1}H NMR and LC-MS. Both diastereoisomers **21** (*dr* 1 : 1)

were characterized by NMR and ESI$^+$-HRMS/MS (see the Supporting Information).

Finally, and as a proof-of-concept, the OSAE technology was applied in an interactive dual Cys-selective modification of Cys-Cys-Bombesin with two maleimides bearing distinct payloads.

Chemistry–A European Journal

Research Article
doi.org/10.1002/chem.202202377

Chemistry Europe
European Chemical
Societies Publishing

After clean *N*-terminal Cys protection with 2-formylphenyl isobutyrate **11**, a maleimide functionalized with a dansyl fluorophore **22** was selectively installed in the in-chain Cys (conjugate **23**). Then, deconjugation of thiazolidine protecting group upon pH acidification released the *N*-terminal Cys which was subsequently functionalized with PEG-maleimide **24** (Scheme 6B). The tunable *N*-terminal Cys protection allowed sequential functionalization of the two Cys to efficiently afford the desired dually modified conjugate **25** as confirmed by LC-MS/MS analysis (see the Supporting Information), in four step reactions performed under dilute and mild conditions.

Conclusion

Preliminary studies show that *O*-ester salicylaldehydes are more efficient handles for the specific protection of *N*-terminal cysteines than other benzaldehydes. Mechanistic and kinetic studies unveiled the relevance of the *O*-ester group in accelerating phenolic thiazolidine formation. Among the different substituted esters, *i*Pr ester showed the best compromise of hydrolytic stability and solubility in aqueous solvents, further ensuring the suitability of OSAE for bioconjugation under dilute and mild conditions. As shown herein, OSAEs allowed the sequential construction of site-selective bombesin peptide dually modified with two distinct payloads orthogonally introduced on Cys residues. Thanks to the high control of stability and cleavage displayed on the conjugation and irreversible deconjugation, OSAEs hold great promise for the design of well-defined constructs with higher levels of complexity, complementing the existing technologies for stable, irreversible and fast-equilibrium systems already reported.

Experimental Section

All detailed experimental procedures and data that support the results reported herein are provided in the Supporting Information.

General procedure for the aldehyde or ketone 1–5, 11–13 and peptide screening: To a 10–25 µM solution of peptide (5.00 nmol) in ammonium acetate solution 20 mM, pH 7.0 was added a solution of tris(2-carboxyethyl)phosphine hydrochloride (TCEP) 3.5 mM in water (15.0 nmol) and the solution mixed for 1–2 h at 23 °C. Then, the aldehyde or ketone **1–5**, **11–13** (10 mM in ACN) (50.0 nmol) was added. After 1 h the reactions were monitored by ESI⁺-MS direct injection and LC-HRMS.

General procedure for mechanistical studies through ¹H NMR: To a 25 mM solution of nucleophile **8-10** (0.014 mmol) in KPi 50 mM in D₂O, pH 7.4 (0.5 mL) was added 2-formylphenyl acetate **1** (0.017 mmol) in [D₆]DMSO (0.05 mL). The reaction was monitored by ¹H NMR spectroscopy.

Procedure for Cys-Cys-Bombesin interactive dual Cys-selective modification

1) *N-terminal Cys protection*: To a 233 µM solution of Cys-Cys-Bombesin (3.50 µmol) in ammonium acetate 20 mM pH 7.0 (14 mL) was added TCEP 3.5 mM in water (10.5 µmol) and the

solution mixed for 1 h at 23 °C. Then, 2-formylphenyl isobutyrate **11** (0.017 mmol) was added in ACN and allowed to react for 1 h for complete *N*-terminal Cys protection. The reaction was monitored by ESI⁺-MS after 1 h to afford the phenolic thiazolidine of Cys-Cys-Bombesin **19**.

2) *In-chain Cys functionalization*: Dansyl-maleimide **22** (10 mM in ACN; 4.20 µmol) was added and allow to react for 1 h, to afford conjugate **23**, as confirmed by ESI⁺-MS and a side product with the installation of two dansyl-maleimides.

3) *Deprotection and 4) N-terminal Cys functionalization*: The pH was adjusted to 4.7 with acetic acid and PEG-Maleimide **24** (0.175 mL, 0.017 mmol) was added immediately after for thiazolidine deprotection and installation of the second maleimide unit. The final construct **25** structure was validated by ESI⁺-MS and MS/MS analysis to the reaction mixture after 2 h.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: bioconjugation · irreversible deconjugation · *N*-terminal cysteine · protecting group

[1] a) C. D. Spicer, E. T. Pashuck, M. M. Stevens, *Chem. Rev.* **2018**, *118*, 7702; b) J. A. Shadish, C. A. DeForest, *Matter* **2020**, *2*, 50; c) C. B. Rosen, M. B. Francis, *Nat. Chem. Biol.* **2017**, *13*, 697; d) E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira, G. J. L. Bernardes, *Nat. Chem. Rev.* **2019**, *3*, 147; e) N. Krall, F. P. da Cruz, O. Boutureira, G. J. L. Bernardes, *Nat. Chem.* **2016**, *8*, 103.

[2] a) S. B. Gunnoo, A. Madder, *ChemBioChem* **2016**, *17*, 529; b) P. Ochtrop, C. P. R. Hackenberger, *Curr. Opin. Chem. Biol.* **2020**, *58*, 28; c) Y. Chen, W. Yang, J. Wu, W. Sun, T.-P. Loh, Y. Jiang, *Org. Lett.* **2020**, *22*, 2038; d) A.

Chemistry–A European Journal

Research Article
doi.org/10.1002/chem.202202377

Chemistry Europe
European Chemical
Societies Publishing

Abbas, B. Xing, T.-P. Loh, *Angew. Chem. Int. Ed.* **2014**, *53*, 7491; e) J. Yu, X. Yang, Y. Sun, Z. Yin, *Angew. Chem. Int. Ed.* **2018**, *57*, 11598.

[3] a) A. Maruani, D. A. Richards, V. Chudasama, *Org. Biomol. Chem.* **2016**, *14*, 6165; b) Y. Seki, T. Ishiyama, D. Sasaki, J. Abe, Y. Sohma, K. Oisaki, M. Kanai, *J. Am. Chem. Soc.* **2016**, *138*, 10798.

[4] a) X. Zheng, Z. Li, W. Gao, X. Meng, X. Li, L. Y. P. Luk, Y. Zhao, Y.-H. Tsai, C. Wu, *J. Am. Chem. Soc.* **2020**, *142*, 5097; b) M. J. S. A. Silva, H. Faustino, J. A. S. Coelho, M. V. Pinto, A. Fernandes, I. Compañón, F. Corzana, G. Gasser, P. M. P. Gois, *Angew. Chem. Int. Ed.* **2021**, *60*, 10850; c) A. Istrate, M. B. Geeson, C. D. Navo, B. B. Sousa, M. C. Marques, R. J. Taylor, T. Journeaux, S. R. Oehler, M. R. Mortensen, M. J. Deery, A. D. Bond, F. Corzana, G. Jiménez-Osés, G. J. L. Bernardes, *J. Am. Chem. Soc.* **2022**, *144*, 10396; d) N. Asiimwe, M. F. Al Mazid, D. P. Murale, Y. K. Kim, J.-S. Lee, *Pept. Sci.* **2022**, *114*, e24235.

[5] R. J. Spears, C. McMahon, V. Chudasama, *Chem. Soc. Rev.* **2021**, *50*, 11098.

[6] a) G. J. L. Bernardes, M. Steiner, I. Hartmann, D. Neri, G. Casi, *Nat. Protoc.* **2013**, *8*, 2079; b) D. Bermejo-Velasco, G. N. Nawale, O. P. Oommen, J. Hilborn, O. P. Varghese, *Chem. Commun.* **2018**, *54*, 12507.

[7] a) H. Faustino, M. J. S. A. Silva, L. F. Veiros, G. J. L. Bernardes, P. M. P. Gois, *Chem. Sci.* **2016**, *7*, 5052; b) A. Bandyopadhyay, S. Cambray, J. Gao, *Chem. Sci.* **2016**, *7*, 4589.

[8] J. P. M. António, H. Faustino, P. M. P. Gois, *Org. Biomol. Chem.* **2021**, *19*, 6221.

[9] a) K. Li, W. Wang, J. Gao, *Angew. Chem. Int. Ed.* **2020**, *59*, 14246; b) D. P. Murale, S. C. Hong, S.-y. Jang, J.-S. Lee, *ChemBioChem* **2018**, *19*, 2545.

[10] R. G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press: New York, **1989**.

[11] DFT calculations at the M06-2X/6-31+G(d,p) level were performed with the *Gaussian 09* package. The model used in the calculations included an explicit water molecule and solvent effects (water) were further considered by means of the PCM model. The Supporting Information includes a complete account of the computational details, the corresponding list of references, the complete reaction free energy profile and the coordinates of all the optimized species.

[12] K.-S. Lee, T.-K. Kim, J. H. Lee, H.-J. Kim, J.-I. Hong, *Chem. Commun.* **2008**, 6173.

[13] N. Naruse, D. Kobayashi, K. Ohkawachi, A. Shigenaga, A. Otaka, *J. Org. Chem.* **2020**, *85*, 1425.