RSC Chemical Biology

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: Y. Chuang, T. Yang, Y. Chih, P. Kuo, C. Hsu, S. Chiou and J. Chu, *RSC Chem. Biol.*, 2026, DOI: 10.1039/D6CB00152A.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/rsc-chembio

View Article Online
DOI: 10.1039/D6CB00152A

Protein Digestion Using a Cysteine-Specific Backbone Cleavage Reagent

Yu-De Chuang[1†], Tsung-Jung Yang[1†], Yu-Chi Chih[1], Po-Cheng Kuo[1], Cheng-Chih-Hsu[1], Shao-Lun Chiou[1*], John Chu[12*]

[1]Department of Chemistry, National Taiwan University, Taipei 106319, Taiwan. [2]Center for Emerging Material and Advanced Devices, National Taiwan University, Taipei 10617, Taiwan

† These authors contributed equally to this work.
* Correspondence: d11223104@ntu.edu.tw and johnchu@ntu.edu.tw

Abstract (197 words)

Mass spectrometry (MS)-based protein analysis is an indispensable tool in modern biomedical research. A key step in sample preparation is proteolytic digestion using enzymes with well-defined amino acid specificity, such as trypsin, chymotrypsin, and StaphV8 protease, which cleave at basic, aromatic, and acidic residues, respectively. The absence of cysteine (Cys)-specific cleavage methods is a gap in the current protein analysis toolbox. Herein, we report a chemical reagent (**1**) that selectively cleaves the N-terminal amide bond of Cys residues in proteins. Using glutathione as a model peptide, we investigated the reaction kinetics in detail and based on which identified an optimized conditions for protein cleavage. Using thioesterase as a model protein, we further demonstrated that **1** is fully compatible with modern MS-based proteomics workflows, including in-gel digestion and use in combination with existing proteases. This reaction proceeds rapidly and selectively in aqueous buffers, affording high yields while converting the reactive Cys side-chain thiol into a chemically inert five-membered heterocyclic moiety. This transformation eliminates the need for the commonly employed iodoacetamide capping step and introduces a distinct mass tag that facilitates downstream data analysis. Overall, these features establish **1** as a robust and practical new tool for protein analysis.

Introduction

Proteins carry out the majority of cellular processes, and mutations that disrupt their functions are often key drivers of diseases. Consequently, methods for protein analysis, including the identification and sequencing of proteins, are central to modern biomedical research.[1] Sequencing used to be accomplished by Edman degradation, an iterative procedure that uses the chemical reagent phenyl isothiocyanate to cleave one amino acid (AA) at a time from the N-terminus of a peptide.[2] The released AA is then identified by comparison with known standards via thin layer chromatography. With the advent of high-resolution mass spectrometry (MS), Edman degradation has been replaced by liquid chromatography (LC) coupled to MS (or tandem MS when needed).[3] The standard workflow in modern proteomics includes two key steps: digestion and sequencing (*Fig. 1a*).[4] *Digestion* involves breaking down a protein into peptide fragments, which are subsequently analyzed by LCMS to determine the sequential order of their constituent AAs.

Methods that cleave a protein only at a select type(s) of AA are particularly useful, and for this reason, proteases with high AA specificity have been central to protein analysis.[5] Trypsin, for example, catalyzes amide bond hydrolysis at the C-terminal side of arginine (Arg) and lysine (Lys) residues and is the most widely used enzyme for this purpose. Chymotrypsin and V8 protease cleave at aromatic and acidic AAs, respectively. Furthermore, digesting the same protein using

View Article Online
DOI: 10.1039/D6CB00152A

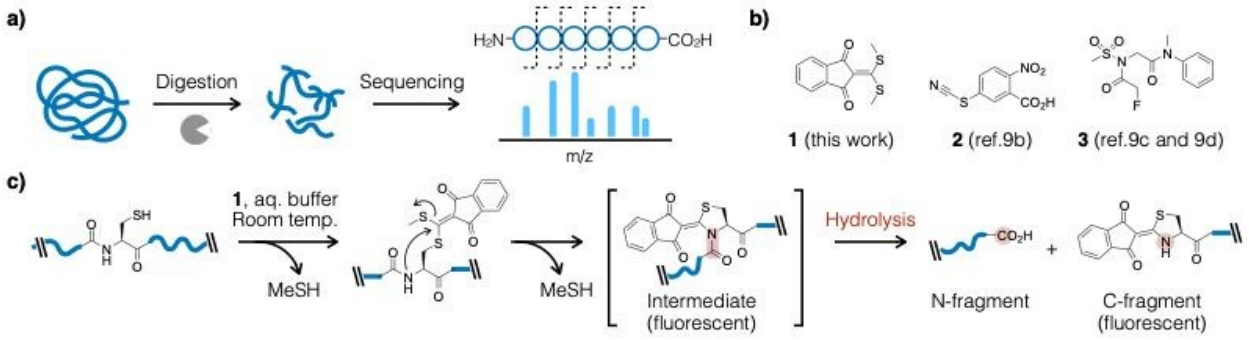

Figure 1. a) Modern proteomics analysis begins with proteolytic digestion. The resulting peptide fragments are then sequenced by MS and tandem MS. **b)** Chemical reagents capable of inducing amide bond cleavage on the N-terminal side of cysteine residues. **c)** Herein, we present **1** as a new reagent for selective backbone cleavage at Cys residues. The reagent first forms a five membered heterocyclic intermediate, and then its N-terminal amide bond undergoes base promoted hydrolysis.

proteases with distinct specificities in separate reactions generates peptide fragments with overlapping sequences. This strategy facilitates de novo protein sequencing, middle-down analysis, or distinguishing homologous proteins.[6]

However, very few proteases exhibit strict AA specificity. Several chemical reagents have been developed to bridge this gap (*Fig. 1b* and *Fig. S1*). For example, cyanogen bromide (CNBr), which cleaves proteins at methionine residues, was once widely used.[7] However, its high toxicity and sensitivity to oxidation are obvious drawbacks, and it is almost completely absent from routine protein analysis today. More recently, Raj and coworkers reported a serine (Ser) specific cleavage reagent; however, the reaction conditions were harsher than most biological samples can tolerate.[8] While no known protease cleaves exclusively at cysteine (Cys), several chemical reagents target Cys by first modifying the side-chain thiol, such as through cyanylation and acetylation by reagents **2** and **3**, respectively (*Fig. 1b*), followed by backbone cleavage at these positions.[9] Unfortunately, the moderate cleavage efficiency of these reagents has limited their widespread adoption in protein analysis. Reported herein is reagent **1**, a chemical tool that enables highly efficient and selective cleavage at Cys residues (*Fig. 1c*).

Using glutathione (GSH) as a model peptide, we showed that **1** cleaves quantitatively the N-terminal amide bond of Cys in aqueous buffer at pH 11. We also used macolacin thioesterase (mTE) as a model protein to showed that **1** is compatible with standard protein analysis workflow. Specifically, **1** can be used alone or in combination with other proteases for in-gel digestion of proteins, and the resulting peptide mixtures are directly amenable to LCMS analysis. An additional advantage of **1** is that it converts the free Cys side-chain thiol into a chemically inert five membered heterocyclic moiety, thereby obviating the commonly used iodoacetamide capping step that prevents unwanted thiol side-reactions.[10] This heterocyclic moiety also serves as a distinct mass tag that facilitates fragment identification during MS analysis. Collectively, these finding establish **1** as a valuable new tool for protein analysis.

Results and Discussion

This work was inspired by a recent publication from the Sun research group.[11] They reported **1** as a protein labelling reagent that reacts selectively with the Cys side-chain thiol, and then condenses with the amide NH to form a conjugated five membered heterocyclic intermediate (**4**, *Fig. 1c*). The N-terminal amide bond of Cys becomes connected to this moiety and starts to

View Article Online
DOI: 10.1039/D6CB00152A

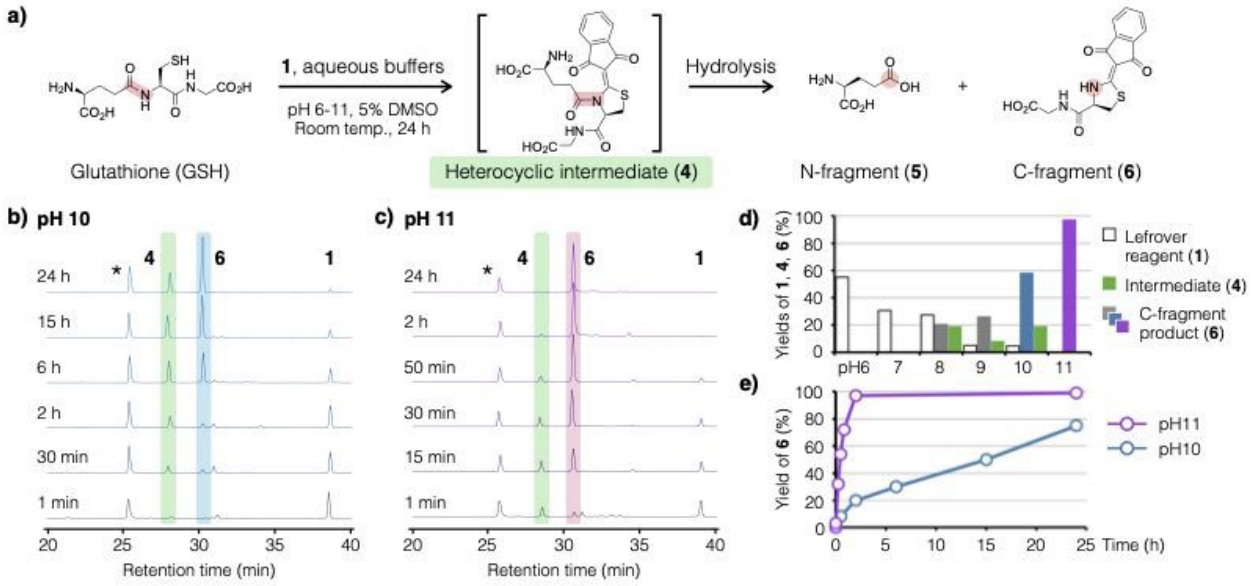

Figure 2. a) Glutathione (GSH) is used as a model peptide to study amide bond cleavage by **1** in aqueous solutions. Reactions at pH 10 **(b)** and pH 11 **(c)** were quenched at various time points and analyzed by HPLC. Caffeine was used as the internal standard (*) to facilitate quantitation. Time courses of the reaction at pH 10 **(d)** and pH11 **(e)** were visualized by showing the consumption of the reagent (**1**, empty bars), accumulation of the intermediate (**4**, green bars), and formation of the C-fragment product (**6**, gray bar show yields at pH 8 and 9; blue and purple bars show yields at pH 10 and 11, respectively)

undergo slow hydrolysis even at near neutral pH. Hydrolysis leads to backbone cleavage and is considered an undesired side reaction for a protein labelling reagent. Nevertheless, the same reaction becomes advantageous if peptide bond cleavage is the intended outcome. With this in mind, we explored the possibility of repurposing **1** as a reagent for Cys specific protein digestion.

We screened for conditions that enhanced amide bond hydrolysis (*Fig. 2a*, *Fig. S1-S5*). Our first series of model reactions was setup using GSH at room temperature in aqueous buffers adjusted to pH 6, 7, 8, 9, 10, and 11, wherein one equivalent of **1** was mixed with GSH and incubated at room temperature (*Table S1* and *Fig. S8*). After 24 hours, the reaction was quenched by adding formic acid to a final concentration of 2% (v/v). Reagent **1,** the heterocyclic intermediate **4**, N-fragment **5**, and C-fragment **6** were all readily separable by HPLC. The reaction yields were determined based on integrating the peak that corresponds to the C-fragment **6**. We monitored the course of this reaction in detail at pH 10 and 11 by removing small aliquots from the reaction mixture at various timepoints. These aliquots were immediately quenched and analyzed by HPLC. The results showed that GSH was approximately 50% hydrolyzed after 24 hours at pH 10 (*Fig. 2b*, *Table S2*, and *Fig. S9*) and was completely hydrolyzed within 2 hours at pH 11 (*Fig. 2c*, *Table S3*, and *Fig. S10*). These data showed that basic conditions promote GSH cleavage by **1**, resulting in both faster rate and higher yields (*Fig. 2d* and *2e*).

A series of kinetic models were constructed to try to describe the course of this reaction (*Fig. S11* to *S16*). Data fitting did not noticeably improve when the reactions were allowed to proceed in reverse. Therefore, we chose a simplified model that describes the addition-elimination that leads up to intermediate **4**, as well as the hydrolysis that give rise to fragments **5** and **6**, both as (nominally) irreversible steps (*Scheme 1*). Then, based on a global fit that allows the reaction order of each component to vary, we concluded that the first step of this reaction is zeroth order to methanethiol (MeSH) and half order to hydroxide (OH^-), which likely stems from the fact that

View Article Online
DOI: 10.1039/D6CB00152A

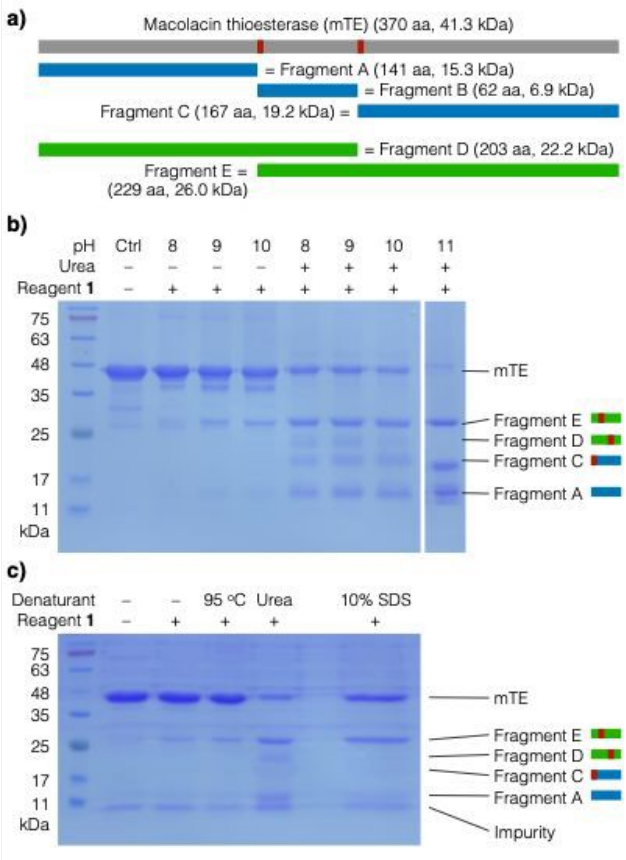

Figure 3. Macolacin thioesterase (mTE) was used as a model to visualize protein cleavage by **1**. **a)** mTE contains two Cys residues (highlighted in red) at position 142 and 204. Cleavage at both Cys yields three possible fragments (A (15.3 kDa), B (7.0 kDa), and C (19.3 kDa)), and cleavage at only one of the two Cys yields two possible fragments (D (22.2 kDa) and E (26.1 kDa). **b)** mTE (10 µM) was mixed with **1** (200 µM) at pH 8 to 11 for 24 h at 25 °C. Fragments A, C, D, and E were readily resolved by SDS PAGE, whereas fragment B, which is too small to be visualized by SDS PAGE, was detected by MALDI-MS (*Fig S21*). The results show that cleavage efficiency is enhanced under basic conditions and in the presence of urea (8 M). **c)** mTE was subjected to various denaturation conditions, including heat (95 °C for 10 min), a chaotropic agent (urea, 8 M), and surfactant (SDS, 10% w/v), and then incubated with **1** for 24 h at pH 10 and 25 °C. The results show that the highest cleavage efficiency of **1** was achieved in the presence of urea.

hydroxide promotes the thiol-thiolate equilibrium of the GSH side-chain.[12] The second step is a normal hydrolysis reaction. With these considerations in mind, the rate constants were obtained by fitting the production (and consumption) of each component measured by HPLC at various timepoints (*Fig. S17*). The rate constants k_1 and k_2 are 1.9 $M^{-1.5}s^{-1}$ and 0.4 $M^{-1}s^{-1}$ at pH 10, and 14.0 $M^{-1.5}s^{-1}$ and 2.0 $M^{-1}s^{-1}$at pH 11, respectively.[13] This model provides a quantitative description of this reaction and helps to explain its strong pH dependence.

We then tested **1** for protein digestion using the macolacin thioesterase (mTE) as a model (*Fig. 3a*).[14] It contains two free Cys at residue 142 and 204. Complete hydrolysis of mTE by **1** would therefore produce three fragments that are 15.3 (A), 7.0 (B), and 19.3 (C) kDa in size. Alternatively, partial digestion of mTE, with hydrolysis occurring at only one of the two sites, would result in two pairs of fragments that are 22.2 (D)/19.3 (C) and 26.1 (E)/15.3 (A) kDa in size. Note that the molecular weight of fragment B, C, and E reported herein include that of the heterocyclic moiety resulting from **1** reacting with Cys. We incubated mTE with **1** in the presence and absence of urea at pH 8, 9, 10 and 11 for 24 hours, and then analyzed the reaction mixture by SDS PAGE (*Fig. 3b*). mTE remained intact when no urea was added, suggesting that denaturation is crucial for protein digestion using **1**. In the presence of urea (8 M), fragments A, B, D, and E were readily observed by SDS PAGE, and the smallest fragment B was identified by MS (*Fig. S18*). Different

Scheme 1. The two-step reaction sequence that cleaves GSH into the N-fragment **5** and the C-fragment **6**.

View Article Online
DOI: 10.1039/D6CB00152A

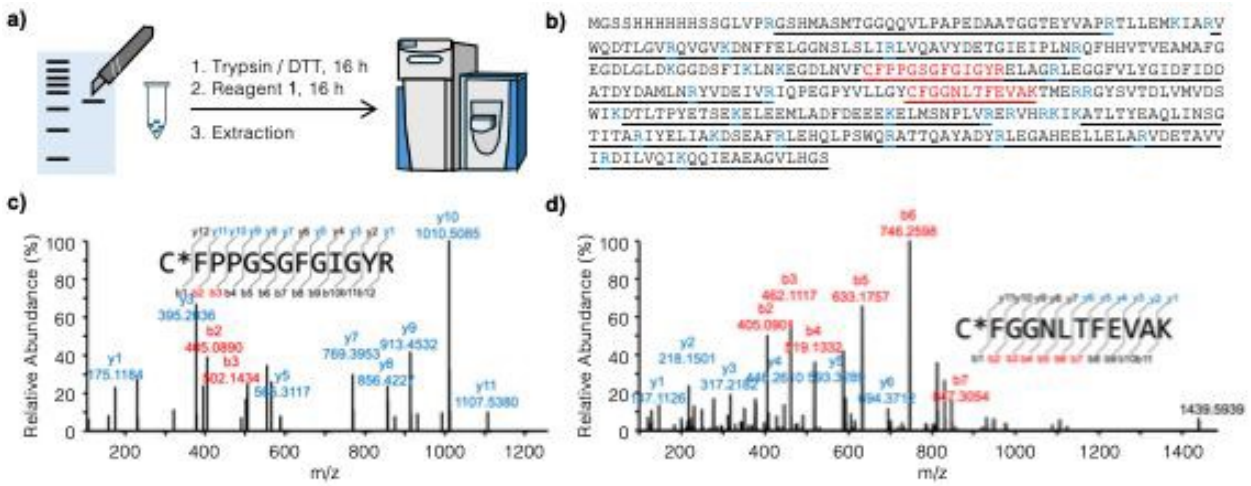

Figure 4. a) The model protein mTE was subjected to the standard protein analysis workflow, including SDS PAGE, in-gel digestion, and LCMS analysis. **b**) Full sequence of mTE is shown. Basic residues (Arg and Lys), which are trypsin cleavage sites, are marked in blue. Cysteine containing fragments are marked in red. MS and tandem MS data of the CFPPGSGFGIGYR fragment (**c**) and the CFGGNLTFEVAK fragment (**d**). The asterisk denotes the thiazolidinone modification (m/z +152); the b and y series ions are shown in blue and red, respectively.

denaturation conditions were then evaluated at pH 10 (8 M urea, 10% (w/v) SDS, and 95 °C incubation for 10 min), and urea turned out to be the most effective (*Fig. 3c*).

In LCMS based protein sequencing, generating fragments at more than one type of AAs is a useful strategy to increase coverage and facilitate data analysis.[3] We therefore applied our reagent in combination with trypsin and, at the same time, evaluated its compatibility with modern proteomics workflow. mTE was analyzed by SDS-PAGE and in-gel digestion according to established protocols (see *Supporting Information*). The band corresponding to mTE was excised, washed, dehydrated by acetonitrile, and digested by trypsin supplemented with dithiothreitol (DTT) for 16 h. Reagent **1** (2 mM) in CABS buffer (pH 11, 5% DMSO) was then added directedly to the mixture and incubated for another 16 h. While iodoacetamide is typically added to cap free Cys thiols and suppress unwanted side reactions, it is unnecessary in our procedure as thiols are converted into an inert five membered heterocyclic moiety upon reaction with **1**. The resulting peptide fragments were extracted from the excised gel band and injected directly into LCMS. Data analysis was performed by MaxQuant v.2.7 with the following settings: S-Oxidation at Met (+16 Da), deamidation at Gln and Asn (+1 Da), and heterocycle formation at Cys (+154 Da) (*Table S5*).[15] As expected, all Cys containing peptide fragments were identified and contained the five membered heterocyclic moiety; no deamidation or N-terminally modified products were detected (*Fig. 4c*, *4d*, and *S19-S21*).

Two additional experiments were performed to assess the stability and specificity of this reagent. In the stability test, **1** was dissolved in an aqueous DMSO solution (5% v/v) and left at 20 °C for one week and showed no detectable degradation (*Fig. S22*). In the specificity test, 10 equivalents of **1** was incubated with ubiquitin at pH 11 for 24 hours and then analyzed by MS. Ubiquitin contains seven Lys and no Cys residues. No Lys modification was detected and only minor N-terminal modification was observed (*Fig. S23*).

View Article Online
DOI: 10.1039/D6CB00152A

Lastly, we estimated the size distribution of peptide fragments generated under various protein cleavage conditions (*Table 1* and *Fig. S24*). Two virtual libraries were compiled based on the UniProt database.[16] One contains 3,000 proteins selected randomly and the other contains 3,000 proteins with at least one Cys residue. Modern MS instruments can readily detect and sequence peptide fragments 7-20 residues in length.[17] Our analysis shows that the distribution of Cys residues across proteins is highly uneven, such that those that do contain Cys would often be cleaved multiple times by **1** and yield 27.1% of fragments in the desirable size window described above. It should be emphasized that our objective is not to replace proteases with **1** in sample preparation, but rather to provide a complementary new tool. As such, although this value is somewhat lower than that observed for trypsin (35.3%) and chymotrypsin (39.7%), the difference is modest and is well within the practically useful range. Furthermore, **1** can also be combined with known proteases to generate orthogonal fragment libraries, thereby improving sequence coverage.[3]

Table 1. Protein cleavage pattern analysis

		Random proteins		Fragment distribution, proteins with ≥ 1 Cys			
		Cut sites (%) [a]	Fragment [b]	< 7 aa (%)	7-20 aa (%)	> 20 aa (%)	Coverage [c]
1	Reagent **1**	2.4 ± 2.3	42.3 ± 83.9	22.8	50.1	27.1	7.9
2	Trypsin (T)	10.8 ± 3.2	9.4 ± 11.2	53.5	11.2	35.3	43.6
3	Chymotrypsin (C)	7.5 ± 2.9	13.9 ± 18.7	40.5	19.8	39.7	34.5
4	T + C	18.3 ± 3.8	5.6 ± 6.5	70.9	2.8	26.2	49.7
5	T + **1**	12.9 ± 3.4	7.9 ± 9.3	58.8	7.3	33.9	48.4
6	C + **1**	9.7 ± 3.7	10.8 ± 14.8	49.2	13.2	37.6	40.6

[ab] Numbers represent average ± standard deviation; [c] Coverage is defined as the proportion of amino acids that end up in fragments 7-20 aa in length after cleavage (= fragment length × frequency ÷ protein length).

Conclusion

Reported herein is a chemical reagent **1** that cleaves proteins selectively at Cys residues in aqueous buffers. It is fully compatible with the standard MS-based proteomics workflow, including in-gel digestion and to be used in combination with proteases. Although reagent **3** has also been used for protein cleavage, a large excess and extended reaction time (1,000 equivalents for 3 days) was required to achieve only moderate yields (~30%). This limitation is likely stems from its instability in aqueous solution. In contrast, **1** is both more stable in solution and more reactive toward Cys, making it a more convenient and practical reagent for protein analysis.

Another advantage is that **1** simplifies the protein analysis workflow by converting reactive Cys thiols into an inert heterocyclic moiety, thereby eliminating the otherwise necessary iodoacetamide capping step. As a small molecule reagent, **1** offers several additional advantages over enzymatic methods and fills an important gap in the current protein analysis toolbox. First, it is readily produced from commercially available starting materials. Second, it is highly stable and can be stored either as a DMSO solution for up to one week or as a solid for extended periods. Third, whereas fragments of the protease itself can interfere with downstream data analysis, the use of **1** simplifies post-cleavage processing as excess reagent can be readily removed by a size exclusion filter.[18] Taken together, these results demonstrate the utility of **1** as a valuable new tool for protein analysis and suggest that it may find broad applications in proteomics.

View Article Online
DOI: 10.1039/D6CB00152A

Acknowledgements

We thank the mass spectrometry research services of the Consortia of Key Technologies at National Taiwan University for technical support. This work was supported by grants from the National Science and Technology Council, Taiwan (NSTC 113-2628-M-002-014-MY4) and National Taiwan University (115L7726).

Conflicts of interest

There are no conflicts to declare.

Data availability

The datasets supporting this article have been uploaded as part of the supplementary information.

References

[1] R. Aebersold, M. Mann, Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.
[2] P. Edman, A method for the determination of amino acid sequence in peptides. *Arch. Biochem.* **1949**, *22*, 475.
[3] M. Mann, R. C. Hendrickson, A. Pandey, Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **2001**, *70*, 437-473.
[4] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1*, 2856-2860.
[5] a) M. L. Huynh, P. Russell, B. Walsh, Tryptic digestion of in-gel proteins for mass spectrometry analysis. *Methods Mol. Biol.* **2009**, *519*, 507-513; b) P. Giansanti, L. Tsiatsiani, T. Y. Low, A. J. R. Heck, Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nat. Protoc.* **2016**, *11*, 993-1006.
[6] a) D. L. Swaney, C. D. Wenger, J. J. Coon, Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J. Proteome Res.* **2010**, *9*, 1323-1329; b) J. R. Wisniewski, M. Mann, Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal. Chem.* **2012**, *84*, 2631-2637.
[7] W. A. Schroeder, J. B. Shelton, J. R. Shelton, An examination of conditions for the cleavage of polypeptide chains with cyanogen bromide: Application to catalase. *Arch. Biochem. Biophys.* **1969**, *130*, 551-555.
[8] H. E. Elashal, M. Raj, Site-selective chemical cleavage of peptide bonds. *Chem. Commun.* **2016**, *52*, 6304-6307.
[9] a) J. Wu, J. T. Watson, Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping. *Anal. Biochem.* **1998**, *258*, 268-276; b) H.-Y. Tang, D. W. Speicher, Identification of alternative products and optimization of 2-nitro-5-thiocyanatobenzoic acid cyanylation and cleavage at cysteine residues. *Anal. Biochem.* **2004**, *334*, 48-61; cN. Zenmyo, et al., A protein cleavage platform based on selective formylation at cysteine residues. *J. Am. Chem. Soc.* **2025**, *147*, 3080-3091; d)Y. Matsumoto, N. Zenmyo, S. Watanabe, K. Sasaki-Tabata, S. Uchinomiya, N. Shindo, A. Ojida, Backbone cleavage of peptides and proteins via cysteine S-fluoroacetylation. *Chem. Commun.* **2025**, *61*, 11625-11628.
[10] a) S. Sechi, B. T. Chait, Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* **1998**, *70*, 5150-5158; b) T. Muller, D. Winter, Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents. *Mol Cell Proteomics* **2017**, *16*, 1173-1187.
[11] H. Zhang, K. Wei, W. Yu, Y. Wu, X. Qian, E. V. Anslyn, X. Sun, Site-specific chemoselective cyclization and fluorogenic modification of protein cysteine residues: from side-chain to backbone. *J. Am. Chem. Soc.* **2025**, *147*, 32818-32829.
[12] K. A. Connors. Chapter 2: Simple Rate Equations. In *Chemical Kinetics: The Study of Reaction Rates in Solution*, VCH Publishing Group: Weinheim, Germany, 1990; pp.24-58.
[13] a) L. Petzold, Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM J. Sci. Stat. Comput.* **1983**, *4*, 136-148; bP. Virtanen, et al., SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Meth.* **2020**, *17*, 261-272.
[14] Z. Wang, B. Koirala, Y. Hernandez, M. Zimmerman, S. Park, D. S. Perlin, S. F. Brady, A naturally inspired antibiotic to target multidrug-resistant pathogens. *Nature* **2022**, *601*, 606-611.

View Article Online
DOI: 10.1039/D6CB00152A

[15] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367-1372.
[16] C. The UniProt, UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **2025**, *53*, D609-D617.
[17] P. B. Pandeswari, V. Sabareesh, Middle-down approach: a choice to sequence and characterize proteins/ proteomes by mass spectrometry. *RSC Adv.* **2019**, *9*, 313-344.
[18] D. Smolin, N. Totsch, J. N. Grad, J. Linders, F. Kaschani, M. Kaiser, M. Kirsch, D. Hoffmann, T. Schrader, Accelerated trypsin autolysis by affinity polymer templates. *RSC Adv.* **2020**, *10*, 28711-28719.

View Article Online
DOI: 10.1039/D6CB00152A

- The data supporting this article have been included as part of the Supplementary Information.