Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2017, **8**, 214

Mechanistic study of CBT-Cys click reaction and its application for identifying bioactive N-terminal cysteine peptides in amniotic fluid†

Zhen Zheng,‡[a] Peiyao Chen,‡[a] Gongyu Li,‡[a] Yunxia Zhu,‡[c] Zhonghua Shi,[d] Yufeng Luo,[a] Chun Zhao,[b] Ziyi Fu,[b] Xianwei Cui,[d] Chenbo Ji,[d] Fuqiang Wang,*[b] Guangming Huang*[a] and Gaolin Liang*[a]

Received 2nd April 2016
Accepted 10th August 2016

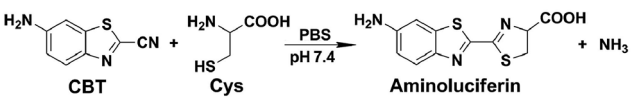
DOI: 10.1039/c6sc01461e

www.rsc.org/chemicalscience

CBT-Cys click condensation reaction has a high second-order reaction rate constant and has found wide applicability in recent years. However, its reaction mechanism has not been experimentally validated and its application for identifying bioactive N-terminal Cys peptides in real clinical samples has not been reported. Herein, firstly, by employing induced nanoelectrospray ionization-mass spectrometry (InESI-MS) and a home-built micro-reactor, we successfully intercepted and structurally characterized the crucial intermediate in this click reaction for the first time. With the intermediate, the proposed mechanism of this reaction was corroborated. Moreover, we also applied this MS setup to monitor the reaction in real time and obtained the second-order reaction rate constants of this reaction at different pH values. After mechanistic study, we applied this click reaction for identifying bioactive N-terminal cysteine peptides in amniotic fluid (AF). Eight unique N-terminal Cys peptides in AF, three of which are located in the functional domain regions of their corresponding proteins, were identified with a false positive rate less than 1%. One of the three peptides was found able to inhibit the growth of uterine endometrial cancer HEC-1-B cells but not the endometrial normal cells *via* a typical apoptotic pathway. With its mechanism satisfactorily elucidated, the kinetic parameters obtained, as well as its application for fishing bioactive N-terminal Cys peptides from vast complex clinical samples, we anticipate that this CBT-Cys click reaction could be applied more widely for the facile isolation, site-specific identification, and quantification of N-terminal Cys-containing peptides in complex biological samples.

Introduction

The click reaction is rapidly becoming an essential tool in medicinal chemistry, combinatorial chemistry, material science, and chemical biology.[1–5] Due to the inherent advantages of a click reaction, including high specificity, quantitative yield, and fidelity under physiological conditions, more efforts have been made on the development and evaluation of more reaction classes possessing these characteristics.[6–8] Recently, based on the proposed regeneration pathway of D-luciferin in firefly, Rao and co-workers developed a novel click reaction with high biocompatibility and controllability.[9] Briefly, as shown in Scheme 1, this click reaction involves the efficient condensation between L-cysteine (Cys) (or D-cysteine) and the cyano group of 2-cyano-6-aminobenzothiazole (CBT) (or 2-cyano-6-hydroxybenzothiazole) to yield **Aminoluciferin** (or luciferin) in D- or L-form. The second-order rate constant of this click condensation reaction was reported to be 26.8 M^{-1} s^{-1},[10] which is 300 times larger than that of the well defined copper-free azide–alkyne cycloaddition (AAC) click reaction (7.6×10^{-2} M^{-1} s^{-1}).[11] To date, this click condensation reaction has been successfully employed to design smart imaging probes (optical, magnetic



Scheme 1 Reaction between 2-cyano-6-aminobenzothiazole (CBT) and cysteine (Cys) to yield the product **Aminoluciferin**.

[a]*CAS Key Laboratory of Soft Matter Chemistry, Department of Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China. E-mail: gmhuang@ustc.edu.cn; gliang@ustc.edu.cn*

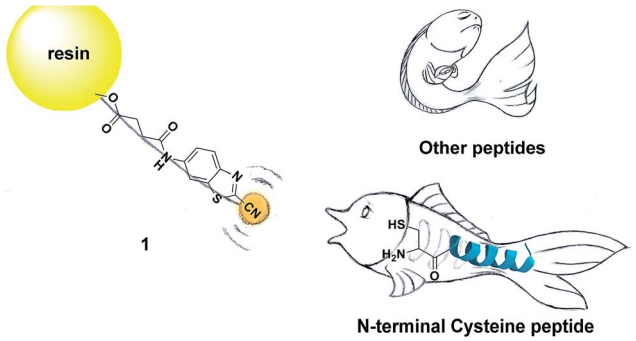
[b]*State Key Laboratory of Reproductive Medicine, Analysis Center, Nanjing Medical University, Nanjing, Jiangsu 210093, China. E-mail: wangfq@njmu.edu.cn*

[c]*Key Laboratory of Human Functional Genomics of Jiangsu Province, Jiangsu Diabetes Center, Nanjing Medical University, Nanjing, Jiangsu, 210093, China*

[d]*Nanjing Maternal and Child Health Institute, Nanjing Maternal and Child Health Care Hospital Affiliated to Nanjing Medical University, Nanjing, Jiangsu 210093, China*

† Electronic supplementary information (ESI) available: General methods; synthesis and characterization; HPLC conditions; Schemes S1 and S2; Fig. S1–S29; Tables S1 and S2. See DOI: 10.1039/c6sc01461e

‡ These authors contributed equally to this work.



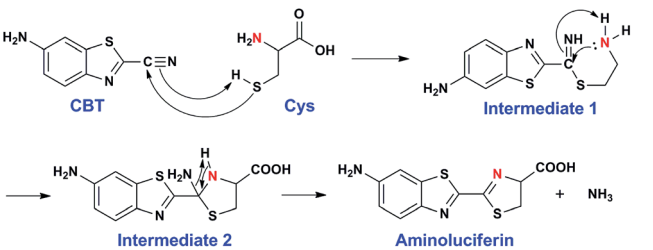
Scheme 2 Schematic illustration of using solid phase CBT to fish N-terminal cysteine (Cys) peptides in the vast complex peptide sample.

resonance, or nuclear),[12–16] synthesize cyclic superstructures,[17] overcome multidrug resistance,[18] and prepare oligomeric hydrogels.[19] Although this CBT-Cys click reaction has been widely used for several years and its reaction equation looks simple, its underlying mechanism has remained less explored. Recently, Liang *et al.* proposed a mechanism for this reaction,[20] however, up to now, there has been no experimental result to validate the mechanism (*e.g.*, direct isolation or characterization of the intermediates in this reaction).

Protein cysteine, serving as an important mediator of redox signaling and regulation, is one of the most reactive amino acid residues in proteins.[21] Remarkably, many of the biologically active peptides contain one or multiple Cys for each, which are believed to contribute to the stability and activity of the peptides.[22–24] However, our overall knowledge of Cys-containing peptides with bioactivity potential is still quite limited owing to the challenges associated with the isolation, site-specific identification, and quantification of these peptides. Conventional methods for the enrichment of Cys-containing peptides are known to use affinity or covalent capture with suitable resins or magnetic beads. For affinity-based methods, affinity capture tags have to be introduced to the Cys-containing peptides through the derivatization of the Cys residues. With the labeled affinity tags, these peptides can be selectively captured by resins (or beads) and then eluted for identification.[25] Thiol–disulfide exchange and thiol–thioester exchange are commonly employed in covalent capture approaches. For example, *via* thiol–disulfide exchange with thiopropyl sepharose on the solid support, Cys-containing peptides can be selectively separated by covalent chromatography.[26,27] Although these methods are able to enrich Cys-containing peptides in a peptide sample pool, the sites of Cys on the enriched peptides (N-terminal, middle, or C-terminal) are not differentiated.

Owing to its capabilities for structural characterization, as well as high specificity, sensitivity, and speed, mass spectrometry (MS) is advantageous over other established spectroscopic techniques which require a conventional periodic sampling–quenching–concentrating–analyzing process for reaction monitoring. Moreover, through interception of reactive intermediates, MS can additionally provide a wealth of mechanistic information.[28–31] Among the MS techniques, electron spray ionization mass spectrometry (ESI-MS) and tandem MS are advantageous in solution-phase and are rapidly becoming the techniques of choice for mechanistic studies in chemistry.[32,33] For example, following the pioneering study by Chen *et al.* on the mechanistic investigation of organometallic chemistry with ESI-MS,[34] De Angelis *et al.* recently uncovered the mechanism underlying the Cu^{+}-catalyzed AAC (CuAAC) click reaction using ESI-MS.[35] Moreover, MS (either ESI-MS or matrix-assisted laser desorption ionization (MALDI) MS) has also been widely used for peptide (or protein) identification.[36,37] Terminal labeling of peptide (or protein) can either simplify the mass spectra or help to distinguish the fragmentation ion series of the peptide (or protein), thus facilitating the direct readout of the amino acid sequence of the peptide (or protein) from the spectra. Therefore, it is of broad interest to develop new approaches to specifically label peptide termini for mass spectrometric analysis.

Inspired by the pioneering studies above, in this work, we firstly aim to investigate the mechanism of CBT-Cys click reaction by using induced nano-electrospray ionization mass spectrometry (InESI-MS), a new emerging MS technique which has shown good performance in matrix tolerance and thus facilitated the direct identification of intermediates in raw reaction mixtures.[38–40] Using a ^{15}N-labeled Cys as one reactant and InESI-MS, we firstly validated that the N-atom in the second thiazole ring of **Aminoluciferin** originated from Cys but not from cyano group of CBT (Scheme 1). Then, we assembled a micro-reactor for InESI-MS analysis and successfully intercepted the crucial **Intermediate 2**, with which the mechanism of this CBT-Cys click reaction was satisfactorily interpreted, as shown in Scheme 3. Finally, a series of CBT-Cys click reactions at different pH values were monitored with our InESI-MS in real time, which gave according kinetic parameters of this reaction. With the mechanism of this CBT-Cys click reaction elucidated, we further employed it for identifying bioactive N-terminal cysteine peptides in amniotic fluid with MALDI-MS. As illustrated in Scheme 2, we rationally designed a CBT derivative, CBT succinic amide (**1**), which has a free carbonic acid group to covalently conjugate to 2-chlorotrityl chloride resin and a light molecular weight (MW) of 275.28 daltons. In the presence of a reducing agent (*e.g.*, dithiothreitol, DTT) and at pH 7.4, the N-terminal Cys peptide in the vast complex peptide sample (*e.g.*, Cys(SEt)-Glu-Tyr(H_2PO_3)-Phe-Phe-Gly-OH, **2** in this work) click reacts with **1** to yield the condensation product which can be cleaved from the resin for mass spectrometric analysis. Employing this strategy and MALDI-MS, we successfully identified eight unique



Scheme 3 Proposed reaction mechanism between CBT and Cys.

N-terminal Cys peptides from human amniotic fluid (AF). Proteomic analyses indicated that three of the eight peptides have C-coil structures and coincidentally are located in the functional domain regions of their corresponding proteins. The cell cytotoxicity study indicated that one of the three peptides showed an obvious inhibitory effect on endometrial cancer HEC-1-B cells. To the best of our knowledge, this is the first time that the N-terminal Cys peptides in real clinic samples are selectively enriched for proteomic analysis.

Results and discussion

Assignment of the origin of the nitrogen atom

We began the study by comparing the InESI-MS and conventional nESI-MS to choose one more suitable means for studying the mechanism of our CBT-Cys click reaction. Fig. S7 and S8† indicated that the molecular ion peak of the product **Aminoluciferin** (*m*/*z* 280) was totally suppressed in conventional nESI-MS due to the interference from the high background peak of the plasticizer (*m*/*z* 278) while that in InESI-MS was clearly observed under the same condition. Direct InESI-MS and corresponding tandem MS, together with high resolution InESI-MS, clearly showed the starting materials, the product **Aminoluciferin**, as well as their characteristic fragment ions (Fig. S9 and S10†). All these results above indicated that InESI-MS had better matrix tolerance and was more suitable than conventional nESI-MS for studying our CBT-Cys click reaction. Therefore, InESI was chosen to investigate the mechanism of CBT-Cys click reaction. As shown in Scheme 1, the CBT-Cys click condensation reaction involves the release of one unit of ammonia gas (NH_3) and the formation of a new thiazole ring (*i.e.*, the second thiazole ring) in the product **Aminoluciferin**. Therefore, whether the N atom in the second thiazole ring of **Aminoluciferin** originated from the cyano group of CBT or the amino group of Cys is critical for resolving the mechanism of this reaction. To assign the origin of this N atom, we employed a stable isotope-labeling method to trace the N atom. In detail, the reaction mixture of CBT with natural Cys or ^{15}N-Cys was separately subjected to InESI-MS and their corresponding MS spectra were recorded for comparative analysis. Fig. 1 showed that **Aminoluciferin** from the reaction mixture of natural Cys and CBT was observed at *m*/*z* 280, while *m*/*z* of **Aminoluciferin** from ^{15}N-Cys and CBT shifted to 281. This suggests that the N-atom in the second thiazole ring of **Aminoluciferin** originated from Cys but not CBT.

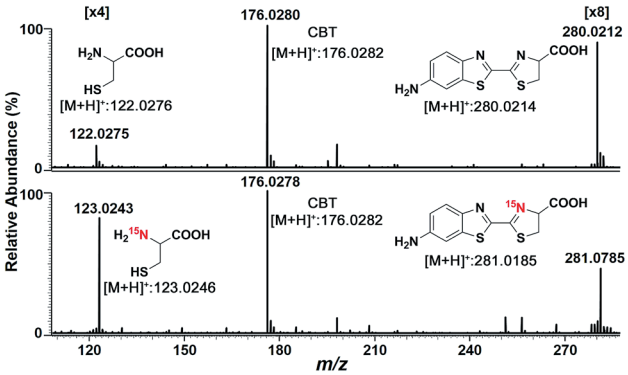


Fig. 1 InESI(+)-MS spectra of the CBT-Cys reaction using natural Cys (top) or ^{15}N-labeled Cys (bottom). Reaction conditions: 50 μM CBT with 50 μM Cys under pH 7.4, $CH_3OH : H_2O = 1 : 1$.

Proposal and validation of the reaction mechanism

After assignment of the N atom in the second thiazole ring of the product **Aminoluciferin**, the mechanism underlying this CBT-Cys click condensation reaction can be roughly drawn, as depicted in Scheme 3. In detail, since the thiol group has a much higher nucleophilicity than the amino group, we propose that the S atom on Cys first attacks the C atom on the cyano group of CBT, causing the N atom on the cyano group to become negatively charged which thereafter attracts a H-atom on the –SH group to yield **Intermediate 1**. Then the amino group on the cysteine attacks the same C atom, while in the meantime the N atom on the enamine of **Intermediate 1** draws a H-atom from the amino group, to form a new thiazolidin structure of **Intermediate 2**. Then the N atom originating from the cyano group of CBT acquires its third hydrogen from thiazolidin to yield NH_3 which is released.

To validate the proposed mechanism above, we need to intercept and characterize the transient intermediates in this click reaction. Obviously, **Intermediate 2** is structurally more stable than **Intermediate 1** but both intermediates have an *m*/*z* 297 (for natural Cys reaction). Structural analysis with the obtained MS data indicated that the intermediate we intercepted in this work was **Intermediate 2**. Taking their very short lifetimes into consideration, interception of the intermediates challenges most of the conventional MS methods. In this study, as shown in Fig. 2A, we introduced a thin fused silica capillary (loaded with the solution of 100 μM CBT) into the nanospray emitter (pre-loaded with the solution of 100 μM Cys) to initiate the reaction in the spray emitter for immediate InESI-MS analysis. The mixing rate of the two reactants was controlled by a syringe pump connected to that fused silica capillary. Once the reactants were mixed, the reaction started and the mixture was immediately subjected to InESI-MS analysis. By this method, the reaction was monitored in real time and transient species were trapped and analyzed. Using this home-built micro-reactor, for the first time, we successfully intercepted **Intermediate 2** and characterized its structure by tandem MS. Fig. 2B showed the selected ion chronograms of reactants Cys (red, *m*/*z* 122†) and CBT (green, *m*/*z* 176), product **Aminoluciferin** (blue, *m*/*z* 280), and the ion *m*/*z* 297 (yellow, the **Intermediate 2**) which recorded their dynamic changes in ion abundance. Note here that the ion chronogram of *m*/*z* 297 was generated simultaneously with the injection of CBT by the springe pump (100 nL min^{-1} for 1 min), suggesting that the intermediate was successfully captured in its rapid occurrence. Further tandem MS analyses structurally resolved that the captured *m*/*z* 297 was the transient **Intermediate 2**. As shown by the MS/MS spectrum in the top panel of Fig. 2C, the ion at *m*/*z* 297 produces a fragment at *m*/*z* 280, which is equal to that of the parent ion of **Aminoluciferin**, as the first prominent product

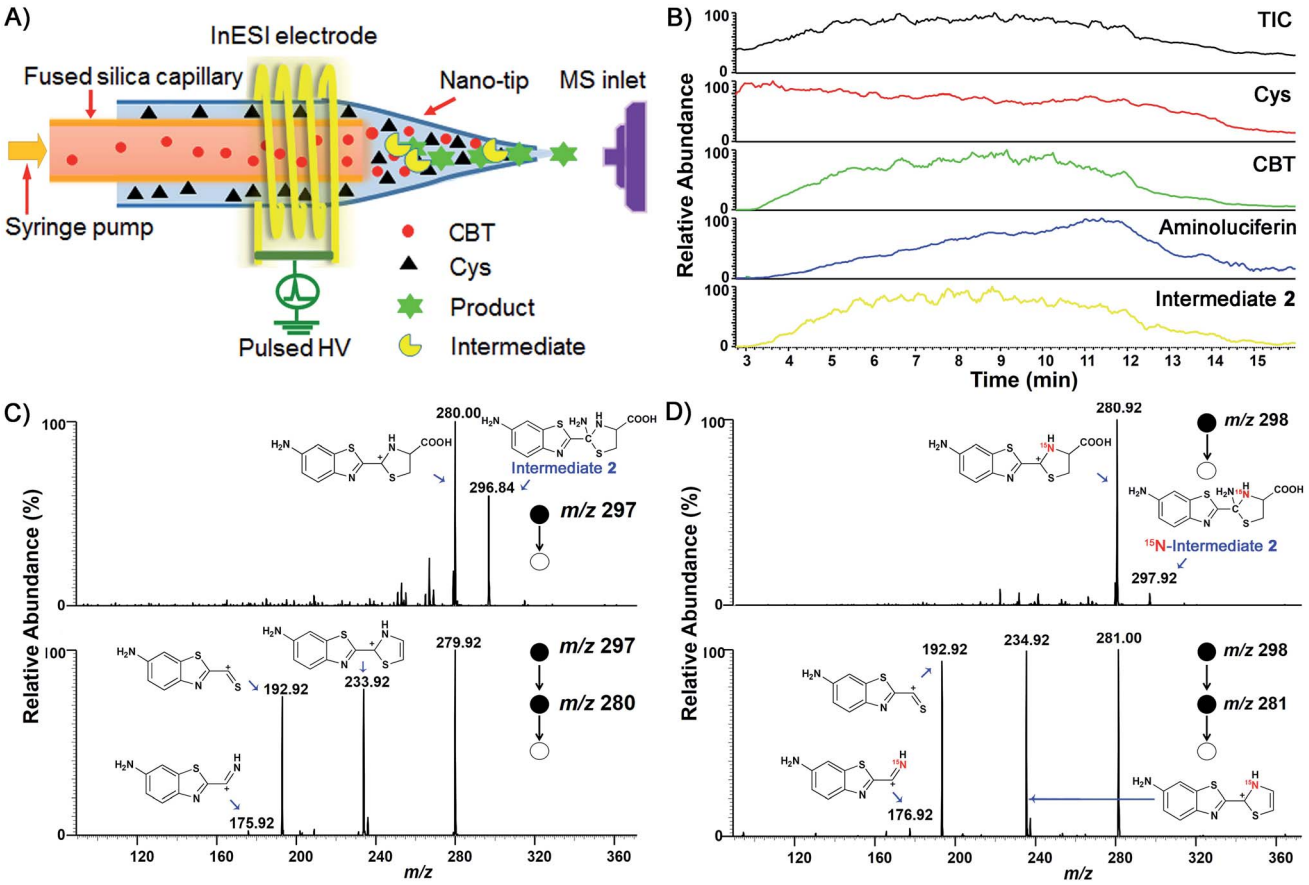


Fig. 2 (A) Instrument illustration to show online CBT-Cys click reaction in the micro-reactor. (B) Ion chronograms of the CBT-Cys click reaction. InESI-(+)-MS/MS (top) and InESI-(+)-MS/MS/MS (bottom) spectra of the intermediates of the CBT-Cys click reaction using natural Cys (C) or ^{15}N-Cys (D) as the reactant. Reaction conditions: pH 7.4, $CH_3OH : H_2O = 1 : 1$.

ion. Further MS^3 fragmentation on the *m/z* 280 ion yielded the ion fragments at *m/z* 234, 193, and 176 (bottom panel in Fig. 2C), which also appeared in the MS^2 spectrum of **Aminoluciferin** (Fig. S9†). These data above suggest that the ion *m/z* 297 is **Intermediate 2** whose structure is depicted in Scheme 3. Similarly, we also used ^{15}N-Cys and CBT for real time InESI-MS analyses and further confirmed the captured intermediate was **Intermediate 2**. As shown in the top panel of Fig. 2D, the ion at *m/z* 298 produces a fragment at *m/z* 281 as the first prominent product ion, which is equal to that of the parent ion of ^{15}N-**Aminoluciferin**. Consistently, further MS^3 fragmentation on the *m/z* 281 ion yielded the fragments at *m/z* 235, 193, and 177 which were corresponding fragment ions of ^{15}N-**Aminoluciferin** (bottom panel in Fig. 2D). These showed that the ion *m/z* 298 is ^{15}N-labeled **Intermediate 2**. Interestingly, the ion fragment *m/z* 193 appeared in both MS^3 spectra of **Intermediate 2** and ^{15}N-labeled **Intermediate 2** because the fragment ion does not contain the N-atom from either natural Cys or ^{15}N-Cys. This additionally verified our structural analysis.

Kinetic study of the CBT-Cys click reaction

After mechanistic study, we also applied our system to investigate the kinetics of this click reaction. The nucleophilicity of the thiol group of Cys is pH-dependent. Therefore, the kinetics of this click condensation reaction at different pH values should be different. To prove this, an online method of uninterrupted analysis of the reaction mixture was designed to monitor the dynamic processes of this reaction in real-time. When the reactants were progressively transformed into the product **Aminoluciferin**, the composition of the reaction mixture was monitored by InESI-MS over time. Since this reaction originally occurs in the firefly body, we firstly studied its kinetics at a physiological condition of pH 7.4. Selected ion chronograms of the reactant Cys (*m/z* 122) and the product **Aminoluciferin** (*m/z* 280) recorded the dynamic changes of their ion abundances, as shown in Fig. S11 (ESI†). As the initial concentration of Cys was known, we thus used the concentration change of Cys but not **Aminoluciferin** for the determination of the second-order reaction rate constant. Applying the similar kinetic analysis reported by Ren *et al.*,[41] we plotted 1/[Cys] *vs.* reaction time and the linear regression analysis of the plot gave the second-order reaction rate constant of 59.7 M^{-1} s^{-1} at pH 7.4, as shown in Fig. 3. This result is consistent in order of magnitude with that recently reported, calculated by the conventional high performance liquid chromatography (HPLC) method (26.8 M^{-1} s^{-1}).[10] Selected ion chronograms of the reactant Cys (*m/z* 122) and the product **Aminoluciferin** (*m/z* 280) in this click

reaction at pH 6 and pH 5 are shown in Fig. S12 and S13 (ESI†). Similarly, their reaction rate constants were calculated to be 9.0 M^{-1} s^{-1} for pH 6 and 2.7 M^{-1} s^{-1} for pH 5, respectively, as shown in Fig. 3. We did not conduct the kinetic study on this click reaction at pH higher than 8 because we found that the reactant CBT was hydrolysed at these conditions.[42,43]

Using solid phase CBT derivative 1 to fish N-terminal Cys peptide 2

After mechanistic study of this CBT-Cys click reaction, we further employed it for identifying bioactive N-terminal cysteine peptides with MALDI-MS. We began the study with the syntheses of compound **1** and peptide **2**. The syntheses are simple and straightforward. Briefly, in the presence of 4-methylmorpholine, CBT reacts with succinic anhydride to yield **1** after high performance liquid chromatography (HPLC) purification (Scheme S1,† Fig. 4A, S1–S3†). Peptide **2** was synthesized with solid phase peptide synthesis (SPPS), followed by the deprotection of the *t*Bu and Boc groups and purified with HPLC (Scheme S2 and Fig. S4–S6†). After syntheses, we conjugated 500 nmol **1** to the resin. Then, different amounts of **2** (from 100 amol to 10 nmol), together with 20 equiv. of DTT, in *N*,*N*-dimethylformamide (DMF) were shaken with the resin overnight at room temperature. After the reaction was completed, the resin was thoroughly washed with DMF and the title product **3** was cleaved from the resin for matrix-assisted laser desorption ionization/time of flight mass spectrometry (MALDI/TOF MS) analyses (Fig. 4A). As shown in Fig. S14 (ESI†), the MS signal of product **3** could be detected with the loading amount of peptide **2** as low as 100 amol. Quantitative analysis indicated that there was a high linearity between the log value of the relative mass signal intensity (SI) of **3** and the substance amount (N) of peptide **2** over the range of 100 amol to 10 nmol ($Y = 1.796 + 0.328X$, $R^2 = 0.990$, Fig. 4B). The recovery test showed that, at the amount of 1–10 nmol of **2**, our method had a recovery rate of 86.8–105.7%, indicating good reliability of the method (Fig. S15 and Table S1†). This suggests that, in combination with MS, our solid phase CBT derivative **1** could be applied to fish and quantitatively detect N-terminal Cys peptides within a wide amount range with very high sensitivity. In addition, both the quantity and sequence identity of the fished peptides were determined by MALDI TOF/TOF MS. MW of the N-terminal Cys peptide was measured by a TOF-MS scan as the parent ion and fragment ions of the parent ion were obtained by the TOF/TOF MS in a LIFT mode.[44] In this work, fished N-terminal Cys peptides were quantified by measuring the relative signal intensities of the peptide's parent ion tagged with the CBT derivative (*i.e.*, **1**) in TOF MS mode, respectively. For example, the standard peptide **2** was fragmented in TOF/TOF MS, and the fragmentation of the b_1 ion did start from the N-terminal disulfided Cys (Cys(SEt)), whose MW is 164.020 (Fig. 4C). Fig. 4D showed that, after reaction with **1** on the resin, the Cys b_1 ion of **2** shifted to 363.034 which corresponds to the b_1 ion of **3**. Therefore in this work, we used *m*/*z* 363.034 as a characteristic fragment ion for qualitative analysis of the N-terminal Cys peptides which conjugated to the "fishhook" (*i.e.*, **1**) on the resin. Combinational analyses of the TOF MS and TOF/TOF MS results of **1**-labeled peptides therefore determined their relative quantities as well as the sequence identities in a single automated operation. Due to the specificity of the CBT-Cys click reaction, every peptide fished by our CBT derivative **1** incontrovertibly starts with an amino acid of Cys from the N-terminal.

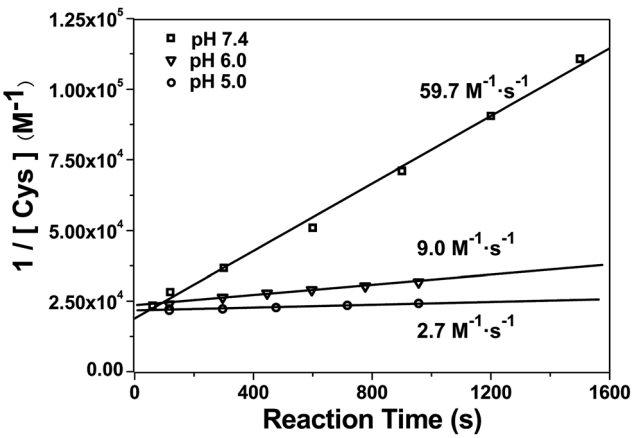


Fig. 3 Linear regression analysis of 1/[Cys] *vs.* time of the CBT-Cys click reaction under different pH values. Reaction conditions: 50 μM CBT and 50 μM Cys, CH_3OH : H_2O = 1 : 1.

Fishing and identification of N-terminal peptides in amniotic fluid

After validation of the feasibility of our method, we applied our solid phase CBT to fish and identify N-terminal peptides in AF for clinical research. AF fills the amniotic cavity that is lined by amnion epithelial cells. The cells are of fetal origin and hence reflect the genotypic constitution of the fetus.[45,46] During pregnancy of the mother, the composition of AF changes, and its protein profile reflects the physiological and pathological changes of the placenta that affect both the fetus and the mother. All the proteins and peptides in the AF, each of which has a MW less than 10 kD, were extracted by centrifugal ultrafiltration and then applied to the CBT resin. The N-terminal Cys peptides reacted with **1** and conjugated to the resin while the unreacted proteins (peptides) were eluted. After being cleaved from the resin and lyophilized, the labeled N-terminal peptides were subjected to MALDI-TOF/TOF MS analysis immediately (Fig. S16†). The acquired MS data of the peptides were analyzed by biotools software Sequence Editor for the determination of their denovo sequences. Then the sequences were searched and validated with software Mascot using the Swiss-Prot human peptide database (none trypsin-digested peptide). In this work, 27 N-terminal Cys peptides in total were "fished" out by our CBT derivative **1**. However, only 17 of them, whose parent ion peaks in the mass spectra showed intensity over 5000 and S/N > 15 to ensure reliable fragment ion peaks, were selected for secondary mass analyses. Based on the parent ions and fragment ions in their MALDI TOF/TOF MS spectra, these 17 peptides were automated denovo sequenced. However, only eight of these 17 N-terminal Cys peptides that were statistically meaningful ($p < 0.05$) were successfully identified (Fig. S17–S24†). Each of

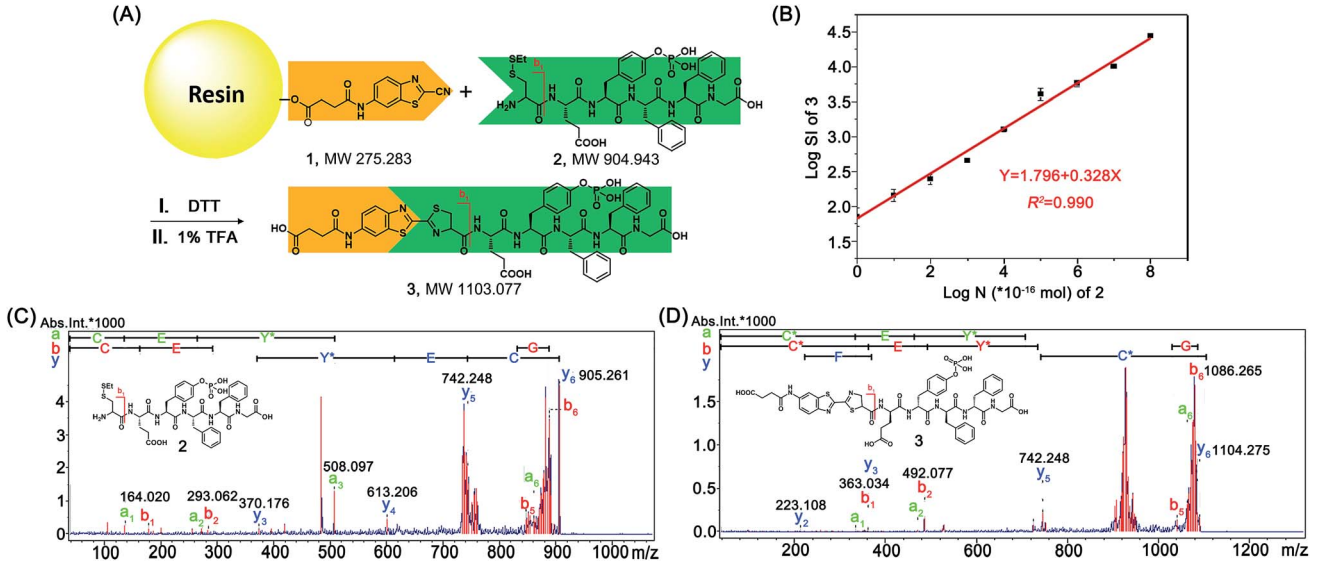


Fig. 4 (A) Schematic illustration of using solid phase CBT derivative **1** to fish N-terminal Cys peptide **2**. (B) The fitted calibration line between the log value of the relative mass signal intensity (SI) of product **3** and the substance amount (N) of peptide **2** ($\times 10^{-16}$ mol). (C) TOF/TOF MS spectrum of **2** showing the pairs of its fragment ions. (D) TOF/TOF MS spectrum of **3** showing the pairs of its fragment ions.

these eight unique free N-terminal Cys-containing peptides has a false positive rate less than 1%, suggesting the effectiveness of our fishing strategy. As we mentioned above, the fragment ions of fished N-terminal peptides should start with the b_1 ion of 363.034. In this work, eight N-terminal Cys peptides were fished and identified, whose peptide sequences, isoelectric points (pIs), and MWs are listed in Table 1.

After fishing and identifying eight N-terminal Cys peptides in AF, we also identified the candidate peptidases that cleave their corresponding proteins to yield the peptides. For example, by searching the peptidase database MEROPS,[47] we identified a number of candidate peptidases that would likely cleave the Ser_{1178}-Cys_{1179} or the Lys_{1121}-Gly_{1122} site of PCSK5_HUAMN (Table S2†). According to Table S2,† matrix metallopeptidase-2 (MMP-2) is the only peptidase that was predicted to be able to cleave both the above two sites of PCSK5_HUAMN to yield the peptide CKTCNGSATLCTSCPK.[48] Therefore, we concluded that the peptide CKTCNGSATLCTSCPK fished by **1** was the cleaved product of PCSK5_HUAMN by MMP-2 (Table S2†). For further function analysis, we predicted the peptide structures using bioinformatic analysis. The sequence-based prediction of the secondary and supersecondary structures of proteins (or peptides) is of great interest to biologists and has found wide applications in numerous areas relating to protein structure and function. Our results indicated that three of the "fished" eight peptides, including peptide CKTCNGSATLCTSCPK, have C-coil structures (Fig. S25†). Coincidentally, using Simple Modular Architecture Research Tool (SMART) and the protein sequence to predict the protein domains of these eight peptide-associating proteins,[49] we found that the above mentioned three peptides were also located in the functional domain regions of their corresponding proteins (Fig. S26†). For example, peptide CQHNTCGGTCDRC was identified in the region of the Laminin EGF-like domain (EGF_Lam) of LAMA5_HUMAN, which contains abundant cysteines to form multiple disulfide bonds.

Table 1 Eight unique free N-terminal Cys-containing peptides identified in AF in this work

Protein	Peptide	Length	pI	MW
K0100_HUMAN	CFLHLP	6	6.73	728.37
MUC19_HUMAN	CGSQCTCQ	8	5.50	828.26
FRITZ_HUMAN	CHQMSFCL	8	6.72	967.37
ZSWM5_HUMAN	CILLLEGGP	9	4.00	913.49
PCSK5_HUMAN	CKTCNGSATLCTSCPK	16	8.53	1615.68
CO4A1_HUMAN	CNGTKCERGPLGPPG	15	8.22	1438.70
LAMA5_HUMAN	CQHNTCGGTCDRC	13	6.71	1396.47
FBN3_HUMAN	CVVPIC	6	5.51	632.30

Identification of a bioactive peptide in the eight "fished" peptides

Since many of the biologically active peptides contain one or multiple Cys, which are believed to contribute to the activity of

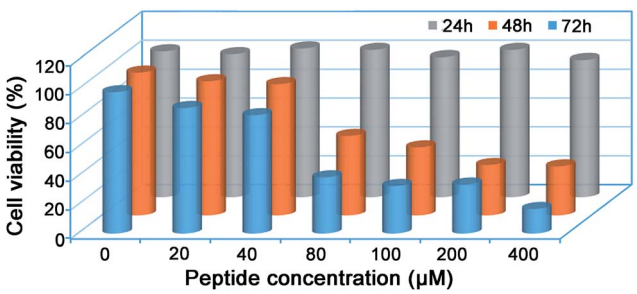


Fig. 5 Cell viability of endometrial cancer HEC-1-B cells treated with peptide TAT-NLS-CKTCNGSATLCTSCPK-NH_2 for 24 h, 48 h, and 72 h, respectively. Each value represents the average value from three independent experiments.

the peptides,[22–24] we proposed that some of our "fished" N-terminal Cys peptides might process biofunctions. Generally, bioactive peptides in AF are defined as those that selectively inhibit the growth of endometrial cancer cells but not normal endometrial cells; we thus investigated the cytotoxicity of the eight "fished" peptides on uterine endometrial cancer cells and normal cells. Before incubating the peptides with cells, we chemically synthesized and modified the peptides by sequentially tagging a transactivator of transcription (TAT, YGRKKRRQRRR) and a nuclear localization signal (NLS, RKRRK) to the N-terminal of each peptide and capped its C-terminal with an amide bond. For example, CKTCNGSATLCTSCPK was modified to be TAT-NLS-CKTCNGSATLCTSCPK-NH_2. TAT is a cell-penetrating peptide from the human immunodeficiency virus, and it can help to deliver proteins, DNA, RNA, or nanoparticles to the cytoplasm in a short time with extremely high efficiency.[50,51] After this modification, each of the eight peptides was assured of being efficiently delivered into the cells. The cell counting kit-8 (CCK-8, Sigma) assay indicated that only the modified peptide TAT-NLS-CKTCNGSATLCTSCPK-NH_2 in the eight peptides inhibited the growth of uterine endometrial cancer HEC-1-B cells, but not the normal endometrial cells such as AA-CELL-24 (Fig. S27 and S28†). In detail, the peptide had half-inhibitory concentrations (IC_{50}) on HEC-1-B cells of 108.6 ± 5.3 μM and 66.0 ± 0.9 μM for 48 h and 72 h, respectively (Fig. 5). To further investigate the inhibitory mechanism of peptide TAT-NLS-CKTCNGSATLCTSCPK-NH_2 on endometrial cancer HEC-1-B cells, after the cells were incubated with 80 μM TAT-NLS-CKTCNGSATLCTSCPK-NH_2 for 24 h and stained with FITC Annexin V and propidium iodide, we applied the fixed cells for flow cytometry analysis. The results indicated that, compared with control cells, peptide-treated cells showed increases of the early-apoptotic stage (Q_3) from 4.5% to 28.0% and late-apoptotic stage (Q_2) from 2.13% to 12.2% (Fig. S29†). But in the necrosis stage (Q_1), the peptide-treated cells (0.055%) did not show an obvious change compared with the control cells (0.031%) (Fig. S29†). All these data above suggest that the peptide TAT-NLS-CKTCNGSATLCTSCPK-NH_2 induces the death of endometrial cancer HEC-1-B cells *via* a typical apoptotic pathway.

Conclusions

In summary, employing InESI-MS and our home-built micro-reactor setup coupled to InESI, we successfully uncovered the mechanism underlying the CBT-Cys click condensation reaction and obtained its second order reaction rate constants at different pH values. Using ^{15}N-Cys as one reactant, we determined that the nitrogen atom in the second thiazole ring of the product **Aminoluciferin** originated from Cys, with which the mechanism underlying this click reaction was proposed. Using our home-built micro-reactor setup coupled to the MS inlet, the crucial **Intermediate 2** in the mechanistic scheme of this reaction was successfully intercepted and structurally characterized for the first time. These InESI-MS and tandem MS evidences corroborated the proposed mechanism. Moreover, using the ion chronograms of the reactant Cys (*m*/*z* 122) to monitor the reaction in real time, we obtained the curves of 1/[Cys] *vs.* reaction time at different pH values. Linear regression analyses of the plots gave the second-order reaction rate constants. For example, the second-order reaction rate constant of this click reaction at pH 7.4 was calculated to be 59.7 M^{-1} s^{-1} with our method, which is consistent in order of magnitude with that recently reported, calculated by the conventional HPLC method (26.8 M^{-1} s^{-1}). After mechanistic study of the CBT-Cys click reaction, we rationally designed a small molecular CBT derivative **1** and developed a facile method of using this solid phase CBT to fish N-terminal Cys peptides in real clinic samples. Using a synthesized N-terminal Cys peptide **2** as the standard and mass spectrometry analysis in a single automated operation, we found our method could be applied to fish and quantitatively detect N-terminal Cys peptides within a wide amount range (100 amol to 10 nmol) with very high sensitivity. Using this method, we identified eight unique N-terminal Cys peptides in AF with a false positive rate less than 1%. Proteomic analysis indicated that three of the eight peptides have C-coil structures and coincidentally are located in the functional domain regions of their corresponding proteins. By tagging a TAT and NLS to each peptide, we studied the cytotoxicity of the eight modified peptides on uterine endometrial cancer cells and normal cells. The CCK-8 assay indicated that only one modified peptide (*i.e.*, TAT-NLS-CKTCNGSATLCTSCPK-NH_2) inhibited the growth of uterine endometrial cancer HEC-1-B cells, but not the normal endometrial cells such as AA-CELL-24. Flow cytometry analysis indicated that the peptide induces the death of endometrial cancer HEC-1-B cells *via* an apoptotic pathway. By selectively isolating N-terminal Cys-containing peptides from a real biological sample, our method significantly reduces the complexity of the peptide mixture in the sample. For example, a theoretical tryptic digest of the entire yeast proteome (6113 proteins) yields 344 855 peptides, among which only 30 619 peptides contain a Cys residue while N-terminal Cys-containing peptides are additionally quite fewer. Thus, if our method is applied for the isolation of the N-terminal Cys peptides in yeast, the complexity of its peptide mixture will be significantly reduced, while its protein quantification and identification can still be achieved. With its mechanism satisfactorily elucidated and kinetic parameters obtained with our InESI-MS method, as well as using this CBT-Cys click reaction to identify bioactive N-terminal Cys peptides in real clinical samples, we anticipate that this CBT-Cys click condensation reaction could be applied more widely for the facile isolation, site-specific identification, and quantification of N-terminal Cys-containing peptides in complex biological samples.

Acknowledgements

This work was supported by Collaborative Innovation Center of Suzhou Nano Science and Technology, the Major program of

Development Foundation of Hefei Center for Physical Science and Technology, Hefei Science Center CAS (2015HSC-UP012), Recruitment Program of Global Experts, and the National Natural Science Foundation of China (Grants U1532144, 31400727, 21375121 and 21475121).

Notes and references

1 A. B. Lowe, C. E. Hoyle and C. N. Bowman, *J. Mater. Chem.*, 2010, **20**, 4745–4750.
2 V. X. Truong, M. P. Ablett, S. M. Richardson, J. A. Hoyland and A. P. Dove, *J. Am. Chem. Soc.*, 2015, **137**, 1618–1622.
3 C. J. Hawker and K. L. Wooley, *Science*, 2005, **309**, 1200–1205.
4 A. Rajendran, M. Endo and H. Sugiyama, *Angew. Chem., Int. Ed.*, 2012, **51**, 874–890.
5 D. J. Ye, A. J. Shuhendler, P. Pandit, K. D. Brewer, S. S. Tee, L. N. Cui, G. Tikhomirov, B. Rutt and J. H. Rao, *Chem. Sci.*, 2014, **5**, 3845–3852.
6 A. E. Speers, G. C. Adam and B. F. Cravatt, *J. Am. Chem. Soc.*, 2003, **125**, 4686–4687.
7 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004–2021.
8 Y. Yuan, D. Li, J. Zhang, X. M. Chen, C. Zhang, Z. L. Ding, L. Wang, X. Q. Zhang, J. H. Yuan, Y. M. Li, Y. B. Kang and G. L. Liang, *Chem. Sci.*, 2015, **6**, 6425–6431.
9 G. Liang, H. Ren and J. Rao, *Nat. Chem.*, 2010, **2**, 54–60.
10 Q. Miao, Q. Li, Q. Yuan, L. Li, Z. Hai, S. Liu and G. Liang, *Anal. Chem.*, 2015, **87**, 3460–3466.
11 J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16793–16797.
12 Y. Yuan, H. Sun, S. Ge, M. Wang, H. Zhao, L. Wang, L. An, J. Zhang, H. Zhang, B. Hu, J. Wang and G. Liang, *ACS Nano*, 2015, **9**, 761–768.
13 A. Dragulescu-Andrasi, S.-R. Kothapalli, G. A. Tikhomirov, J. Rao and S. S. Gambhir, *J. Am. Chem. Soc.*, 2013, **135**, 11015–11022.
14 G. Liang, J. Ronald, Y. Chen, D. Ye, P. Pandit, M. L. Ma, B. Rutt and J. Rao, *Angew. Chem., Int. Ed.*, 2011, **50**, 6283–6286.
15 X. Z. Ai, C. J. H. Ho, J. Aw, A. B. E. Attia, J. Mu, Y. Wang, X. Y. Wang, Y. Wang, X. G. Liu, H. B. Chen, M. Y. Gao, X. Y. Chen, E. K. L. Yeow, G. Liu, M. Olivo and B. G. Xing, *Nat. Commun.*, 2016, **7**, 10432.
16 Y. Liu, Q. Miao, P. Zou, L. Liu, X. Wang, L. An, X. Zhang, X. Qian, S. Luo and G. Liang, *Theranostics*, 2015, **5**, 1058–1067.
17 Y. Yuan, S. C. Ge, H. B. Sun, X. J. Dong, H. X. Zhao, L. N. An, J. Zhang, J. F. Wang, B. Hu and G. L. Liang, *ACS Nano*, 2015, **9**, 5117–5124.
18 Y. Yuan, L. Wang, W. Du, Z. Ding, J. Zhang, T. Han, L. An, H. Zhang and G. Liang, *Angew. Chem., Int. Ed.*, 2015, **54**, 9700–9704.
19 S. Liu, A. Tang, M. Xie, Y. Zhao, J. Jiang and G. Liang, *Angew. Chem., Int. Ed.*, 2015, **54**, 3639–3642.
20 Y. Yuan and G. Liang, *Org. Biomol. Chem.*, 2014, **12**, 865–871.
21 Y. Sato and K. Inaba, *FEBS J.*, 2012, **279**, 2262–2271.
22 H. Antelmann and J. D. Helmann, *Antioxid. Redox Signaling*, 2011, **14**, 1049–1063.
23 N. Brandes, S. Schmitt and U. Jakob, *Antioxid. Redox Signaling*, 2009, **11**, 997–1014.
24 E. Jortzik, L. Wang and K. Becker, *Antioxid. Redox Signaling*, 2012, **17**, 657–673.
25 D. Ren, S. Julka, H. D. Inerowicz and F. E. Regnier, *Anal. Chem.*, 2004, **76**, 4522–4530.
26 J. Paulech, N. Solis, A. V. Edwards, M. Puckeridge, M. Y. White and S. J. Cordwell, *Anal. Chem.*, 2013, **85**, 3774–3780.
27 Y. T. Yao, J. F. Huang, K. Cheng, Y. B. Pan, H. Q. Qin, M. L. Ye and H. F. Zou, *Anal. Chem.*, 2015, **87**, 11353–11360.
28 L. S. Santos, *Eur. J. Org. Chem.*, 2008, **2008**, 235–253.
29 K. L. Vikse, Z. Ahmadi, C. C. Manning, D. A. Harrington and J. S. McIndoe, *Angew. Chem., Int. Ed.*, 2011, **50**, 8304–8306.
30 T. A. Brown, H. Chen and R. N. Zare, *Angew. Chem., Int. Ed.*, 2015, **54**, 11183–11185.
31 T. A. Brown, H. Chen and R. N. Zare, *J. Am. Chem. Soc.*, 2015, **137**, 7274–7277.
32 X. Yan, E. Sokol, X. Li, G. Li, S. Xu and R. G. Cooks, *Angew. Chem., Int. Ed.*, 2014, **53**, 5931–5935.
33 C. Iacobucci, S. Reale and F. De Angelis, *Angew. Chem., Int. Ed.*, 2016, **55**, 2980–2993.
34 C. Hinderling, C. Adlhart and P. Chen, *Angew. Chem., Int. Ed.*, 1998, **37**, 2685–2689.
35 C. Iacobucci, S. Reale, J. F. Gal and F. De Angelis, *Angew. Chem., Int. Ed.*, 2015, **54**, 3065–3068.
36 I. van den Broek, M. Blokland, M. A. Nessen and S. Sterk, *Mass Spectrom. Rev.*, 2015, **34**, 571–594.
37 J. Mayne, Z. Ning, X. Zhang, A. E. Starr, R. Chen, S. Deeke, C.-K. Chiang, B. Xu, M. Wen and K. Cheng, *Anal. Chem.*, 2016, **88**, 95–121.
38 G. M. Huang, G. T. Li, J. Ducan, Z. Ouyang and R. G. Cooks, *Angew. Chem., Int. Ed.*, 2011, **50**, 2503–2506.
39 G. M. Huang, G. T. Li and R. G. Cooks, *Angew. Chem., Int. Ed.*, 2011, **50**, 9907–9910.
40 G. Y. Li and G. M. Huang, *J. Mass Spectrom.*, 2014, **49**, 639–645.
41 H. Ren, F. Xiao, K. Zhan, Y. P. Kim, H. Xie, Z. Xia and J. Rao, *Angew. Chem., Int. Ed.*, 2009, **48**, 9658–9662.
42 J. Jeon, B. Shen, L. Xiong, Z. Miao, K. H. Lee, J. Rao and F. T. Chin, *Bioconjugate Chem.*, 2012, **23**, 1902–1908.
43 H. Chabane, C. Lamazzi, V. Thiery, A. Pierre, S. Leonce, B. Pfeiffer, P. Renard, G. Guillaumet and T. Besson, *J. Enzyme Inhib. Med. Chem.*, 2003, **18**, 167–174.
44 D. Suckau, A. Resemann, M. Schuerenberg, P. Hufnagel, J. Franzen and A. Holle, *Anal. Bioanal. Chem.*, 2003, **376**, 952–965.
45 F. Vesce, B. Pavan, L. Lunghi, G. Giovannini, C. Scapoli, A. Piffanelli and C. Biondi, *Obstet. Gynecol.*, 2004, **103**, 108–113.
46 M. Orczyk-Pawiłowicz, J. Floriański, J. Zalewski and I. Kątnik-Prastowska, *Glycoconjugate J.*, 2005, **22**, 433–442.

47 N. D. Rawlings, A. J. Barrett and A. Bateman, *Nucleic Acids Res.*, 2012, **40**, 343–350.
48 Y.-G. Kim, A. M. Lone, W. M. Nolte and A. Saghatelian, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 8523–8527.
49 I. Letunic, T. Doerks and P. Bork, *Nucleic Acids Res.*, 2015, **43**, D257–D260.
50 H. Nagahara, A. M. Vocero-Akbani, E. L. Snyder, A. Ho, D. G. Latham, N. A. Lissy, M. Becker-Hapak, S. A. Ezhevsky and S. F. Dowdy, *Nat. Med.*, 1998, **4**, 1449–1452.
51 J. M. Gump and S. F. Dowdy, *Trends Mol. Med.*, 2007, **13**, 443–448.