

Azole reagents enabled ligation of peptide acyl pyrazoles for chemical protein synthesis

Peisi Liao^[a] and Chunmao He^{[a]*}

[a] P. Liao, Prof. Dr. C. He

School of Chemistry and Chemical Engineering, South China University of Technology, Guangzhou 510640, China

E-mail: hecm@scut.edu.cn

Abstract

Native chemical ligation (NCL) has been playing an increasingly important role in chemical protein synthesis (CPS), more efficient ligation methods that circumvent the requirement of peptidyl thioester and thiol additive—which allow the following desulfurization or refolding in one pot—are urgently needed for the synthesis of more complex protein targets and in large quantities. Herein, we discover that the weak acyl donor peptidyl *N*-acylpyrazole can be activated by azole reagents like 3-methylpyrazole or imidazole to facilitate its ligation directly with an N-terminal cysteine peptide. As it requires no thioester and thiol additive, this ligation strategy can be conveniently combined with metal-free desulfurization (MFD) or oxidative protein folding to allow various one-pot protocols. The utility and generality of the strategy are showcased by the total synthesis of ubiquitin via an N-to-C sequential ligation-MFD strategy, the semi-synthesis of a copper protein azurin, and the efficient assembly of a sulfated hirudin variant and the cyclotide kalata B1, all in a one-pot fashion.

Introduction

The chemical protein synthesis (CPS), enabled by chemo-selective peptide ligation reactions like native chemical ligation (NCL), has contributed enormously to the generation of proteins with specific structural modifications and facilitated selective introduction of unnatural amino acids, biophysical probes, post-translational modifications.^[1] Since its inception by Kent and co-workers,^[2] and with great contributions from many research groups, NCL has become the most powerful and dominant method for CPS.^[3] With the increasing demand for large and complex proteins, the reactivity of peptidyl thioesters (e.g., alkyl thioesters)^[4] is usually too low to enable NCL under challenging conditions, and thiol additives—with 4-mercaptophenylacetic acid (MPAA) being the most widely employed and gold standard catalyst—have been added to generate more reactive acyl donors to facilitate the ligation.^[3b, 5] A significant advancement in CPS is the development of metal-free desulfurization (MFD) chemistry, which broadened the scope of NCL by allowing ligation sites beyond Cys.^[6] However, the use of MPAA—due to its inherent free radical quenching property^[7]—is problematic when NCL needs to be combined with MFD in one-pot, which is a highly desirable approach to achieve high synthetic efficiency. In this context, a number of strategies have been developed to enable the one-pot NCL-MFD operation, including removal of MPAA through extensive extraction/dialysis and solid-phase

techniques.^[8] Recently, several thiol reagents (poor radical scavengers) have been used to replace MPAA in facilitating one-pot NCL-MFD protocol, including 2-mercaptoethanesulfonate sodium salt (MESNa), trifluoroethanethiol (TFET), methyl thioglycolate (MTG), 2-sulfanylmethyl-4-dimethylaminopyridine (SMDMAP) and thiocholine (**Figure 1A-i**), (see **Table S1** for detailed information).^[9] It is worth noting that sequential C-to-N or N-to-C ligation of multiple segments (≥ 3) is essential in the efficient synthesis of complex proteins; these thiol-additives, however, may complicate the kinetically controlled NCL reactions where two or more peptide thioester segments with different reactivities are involved.^[3b, 9b, 10] Alternatively, non-thiol reagents such as imidazole^[11] and 2-methyl imidazole^[12] have also been shown to serve the purpose (**Figure 1A-ii**). These aforementioned additive strategies, however, require the pre-formed peptidyl thioesters or their surrogates, where although many excellent synthetic methods are available,^[3b, 4c, 13] operational issues and limitations still present, such as relatively complicated post-SPPS manipulations, insolubility of protected peptides, use of large amounts of oxidants and cryogenic conditions and instability of the peptide thioester during storage, etc. Thus, ligation methods that do not require thioesters and thiol additives would be highly efficient in streamlining one-pot NCL-MFD and thus the protein synthesis. To this end, another azole additive (i.e., 1,2,4-triazole) has been employed to assist the one-pot NCL-MFD of peptidyl *N*-acyl-*N*-methyl-benzimidazolinone (MeNbz) (**Figure 1B-i**).^[14] It is worth mentioning that strained peptidyl β -thiolactones are versatile acyl donors allowing the ligation-MFD at Val/Pro-Xaa sites, where no additive is needed for the ligation.^[6g, 15] Unfortunately, these peptide thioester surrogates can only be obtained via Fmoc-SPPS but not recombinantly, which may limit their application in the synthesis of complex and large proteins.

Meanwhile, peptide hydrazide has been developed as an extremely powerful peptidyl thioester surrogate,^[13j, 16] due to its facile access through Fmoc-SPPS or recombinant expression, stability, and convenient thioester transformation.^[17] Recently, Dawson et al.^[18] developed a mild and efficient condensation activation method for peptide hydrazide, where the weakly activated peptide *N*-acyl pyrazoles—generated with stoichiometric amount of acetylacetone (acac)—can be readily exchanged to peptide thioesters in situ, preferably with MPAA, for NCL. Importantly, as peptide thioesters are generated at the addition of stoichiometric quantities of acac, multiple ligations can be allowed sequentially in one-pot through this mild activation strategy. We thus envision that if MPAA can be replaced by thiol-free additives compatible with MFD, it would enable (sequential) ligations at other sites than Cys in one pot, and together with the no requirement of thioester, it should help to greatly simplify purification procedure and increase synthetic efficiency.

Herein, we report that azole reagents (i.e., imidazole and 3-methylpyrazole) are efficient non-thiol additives facilitating the ligation between peptide *N*-acyl pyrazoles with another peptide bearing N-terminal Cys, and we term this strategy *N*-acyl pyrazoles and cysteine ligation (APCL) (**Figure 1B-ii**). While peptide *N*-acyl pyrazoles can be readily obtained through addition of acac, no further exchange into thioester is needed; rather it can react directly with N-terminal Cys peptides with the help of imidazole or 3-methylpyrazole, forming native

peptide bonds. As such, APCL is well compatible with following one-pot MFD or protein folding operations. The utility and generality of the strategy are demonstrated through the total synthesis of ubiquitin via an N-to-C sequential ligations-MFD strategy, the semi-synthesis of a copper protein azurin via a ligation-folding strategy, and the efficient synthesis of a sulfated hirudin (sY⁶³-HV1) protein and the cyclotide kalata B1 (kB1), all in a one-pot fashion.

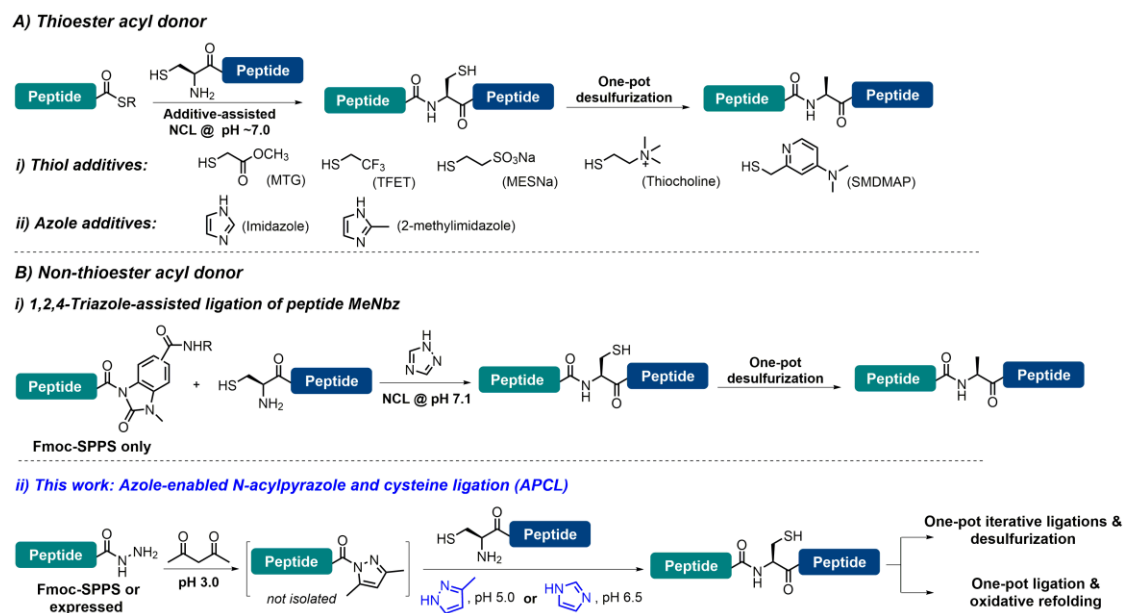


Figure 1. (A) Selected examples of one-pot NCL-MFD strategy using C-terminal α -thioester as the acyl donor. (B) (i) One-pot NCL-MFD strategy using C-terminal MeNbz as the acyl donor; (ii) The principles of the azole reagents enabled APCL strategy and its following one-pot operations described in this work.

Results and Discussion

Condition screening of the APCL strategy

The weakly activated peptide *N*-acyl pyrazole can be readily transformed into peptide thioester/selenoester for ligations,^[18-19] and initially we were interested to see whether its ligation to another Cys-peptide could occur without any additive. Our study began with the ligation between model peptides H-CTFPGASALMKGTLTK-NH₂ (**1**) and Fmoc-RPGKVTEQA-NHNH₂ (**2**). The ligation involves two steps, peptide hydrazide **2** was firstly converted to the corresponding *N*-acyl pyrazole **2a** at pH 3.0, Cys-peptide **1** was then added and the ligation was carried out at various pH ranged between 3.0 and 6.5 for 3 h, before the yield was determined by HPLC. It was observed that the ligation yield was pH-dependent and surprisingly high at pH 6.5 (41%) (Table S2, entries 1–4), which encouraged us to further improve the reaction. To this end, *N*-acyl heterocycles such as *N*-acyl imidazole are well known strong acylating agents,^[20] it was thus hoped that azole reagents would replace the pyrazole moiety to activate the peptidyl acyl group^[21] and prompt its ligation with Cys-peptide directly.

A series of azole reagents including imidazole, 1,2,4-triazole, pyrazole, 3-methylpyrazole and 3,5-dimethylpyrazole were screened for their ability in facilitating the ligation at pH 5.0 and 6.5, which are relatively low compared to traditional NCL in order to minimize possible hydrolysis of **2** while maintaining the nucleophilicity of N-terminal Cys of peptide **1** at a reasonably high level. As shown in **Figure 2** and **Table S2**, we were delighted to observe that ligations performed in the presence of 2.5 M of 3-methylpyrazole at pH 5.0 and 2.5 M of imidazole at pH 6.5 both gave very high HPLC yield after 3 h, rivalling the gold standard catalyst MPAA (compare **Table S2, entry 13 vs 5; entry 16 vs 6**). Further, the effect of the additive concentration was also probed and both additives worked best at 2.5 M (**Figures S27–28**), agreeing well with literatures.^[5b, 11a, 12, 14] It is worth to mention that both additives are inexpensive and highly water-soluble, which elute early and will not affect the HPLC trace. Importantly, the epimerization issue was studied for both additives, and no epimerization at the ligation site was observed during the process (**Figure S31**). Further control ligations were carried out between peptide **2** and peptide **1a** with Ac_m-protected N-terminal Cys or peptide **1b** with N-terminal Ser, and only complete hydrolysis of peptide **2** could be observed in both cases (**Table S3, Figures S32–35**). It suggests that, instead of direct aminolysis, the azole enabled APCL proceeds through a thiol capture and S→N acyl shift reaction pathway similar to NCL, which ensures the chemo- and stereo-selectivity of the ligation.

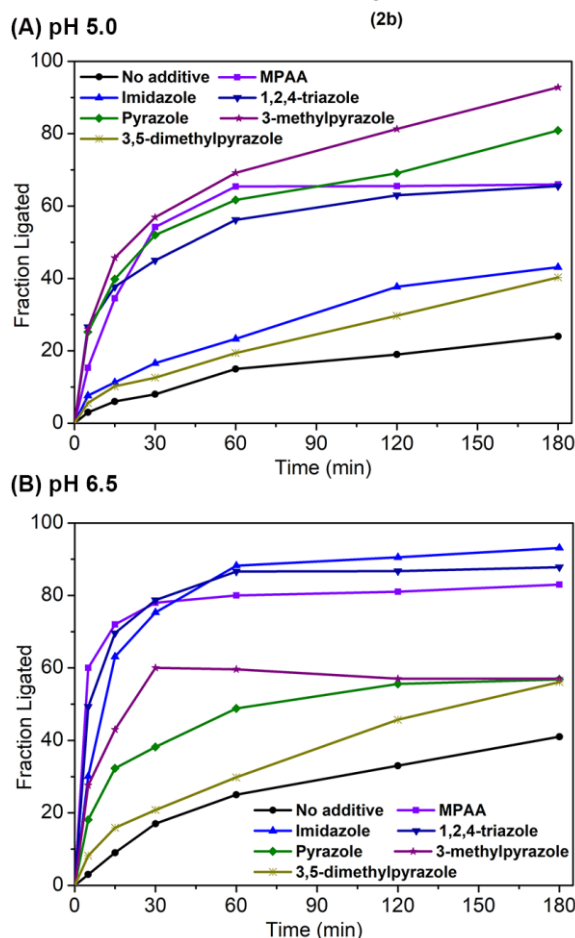
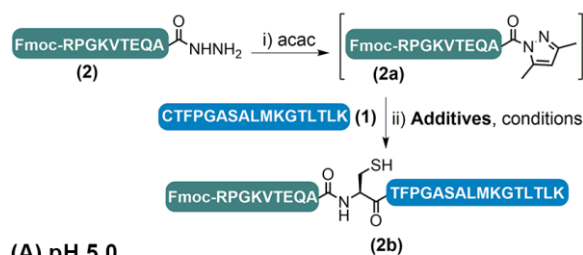


Figure 2. Time-course analysis of APCL with different additives at pH 5.0 (A) and 6.5 (B). Conditions: i) 6 M Gdn-HCl, 0.2 M Na₂HPO₄, pH 3.0, 2.5 eq of acac, 37 °C, 40 min; ii) 6 M Gdn-HCl, 0.2 M Na₂HPO₄, 20 mM TCEP, 2.5 M additives (or 100 mM MPA), 37 °C, pH 5.0 or 6.5, 3 h.

The scope of the APCL strategy

Having determined that the APCL strategy could be used as a viable method for peptide ligation, we proceeded to investigate the scope of the strategy by varying the C-terminal residue of the hydrazide segments (**Table 1**, **Table S4**). The initial acac activation was carried out for 15–150 minutes at pH 3.0, and the following ligation at most C-terminal residues tested could be completed within 3 h with decent isolated yields (47–69%), except for the sterically demanding residues like Val and Pro. Ligation at Glu was complicated by side chain cyclization and isomerization and other known problematic sites like Asp and Asn for obtaining the corresponding peptide hydrazides were not included in the test.^[13j, 18] Notably, in the case of C-terminal Lys peptide **9**, although an excellent isolated yield of 67% was obtained when using 3-methylpyrazole as the

additive, a noticeable amount of side-chain lactamization side product (**9'** in **Figure S42**) could be observed,^[22] which may be sequence specific as this side-reaction was not so obvious in other peptides bearing C-terminal Lys residues (vide infra). In addition, similar level of lactamization could be observed when conducting the ligation with the conventional NaNO₂-MPAA method (data not shown). Lastly, sensitive residues like C-terminal Cys and Trp and internal Cys are well tolerated. Collectively, these results demonstrate the practicability of the azole reagents enabled APCL strategy for peptide ligation.

Table 1. Scope of the ligation of peptide hydrazides.^[a,b,c]

Entry	Ent	Ligation site	Ligation time	Isolated yield
	(X)	(h)	(%) ^[d]	
1	Ala (3)	3	48 ^[b]	
2	Leu (4)	3	47 ^[b]	
3	Gly (5)	3	57 ^[b]	
4	Trp (6)	4	69 ^[b]	
5	Cys (7)	3	48 ^[b]	
6	Phe (8)	3	61 ^[b]	
7	Lys (9)	2	67 ^[b]	
8	Pro (10)	15	14 ^[c]	
9	Val (11)	6	26 ^[b]	
10	Glu (12)	2	5 ^[b]	
11	Gln (13)	2	54 ^[c]	
12	-- (14) ^[e]	2	65 ^[b]	

[a] Conditions: i) 6 M Gdn-HCl, 0.2 M Na₂HPO₄, 2.5 eq of acac, 37 °C, pH 3.0; ii) 6 M Gdn-HCl, 0.2 M Na₂HPO₄, 20 mM TCEP, 2.5 M azoles, pH 5 or 6.5, 37 °C.

[b] 3-methylpyrazole additive, pH 5.0. [c] Imidazole additive, pH 6.5. [d] The weighed isolated yield of the ligation product. [e] Sequence of the model peptide hydrazide **14**: Fmoc-GCGYLKFKSLA-NHNH₂.

One-pot ligation–MFD chemistry

With the knowledge that non-thiol azole additives may not interfere with any subsequent MFD, we next carried out model peptide tests to investigate their suitability for one-pot ligation–MFD chemistry. In initial screenings, unwanted degradation (data not shown) of the peptides occurred at above pH 6.5 (required for

MFD) in the presence of the 3-methylpyrazole but not imidazole, hence it was decided that imidazole could be a more suitable catalyst compatible with one-pot ligation–MFD protocol. In this case, the ligated mixture of **2b** (containing 2.5 M imidazole) was subjected directly to MFD (see SI for experimental details), where additional TCEP (Tris(2-carboxyethyl)phosphine hydrochloride) (final conc. in 200 mM), the radical initiator VA-044 (2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride) (25 eq) and the hydrogen source *t*-BuSH (tert-butylthiol) (5%, v/v) were added and incubated at 37 °C for 8 h to ensure complete desulfurization. Gratifyingly, the one-pot ligation–MFD proceeded smoothly in the presence of imidazole, giving the desired product **2c** in 85% HPLC yield (**Figure S50**).

Total synthesis of ubiquitin through one-pot sequential ligations and MFD strategy

As previously reported peptide hydrazides could be converted into acyl pyrazoles efficiently with stoichiometric amount of activating agent,^[18] which suggested a potential to use this chemistry, powered by imidazole catalysis, to perform multiple sequential ligations and MFD without intermediate purification steps. We attempted to apply this strategy to the synthesis of ubiquitin, which is an important and commonly used model protein consisting of 76 amino acid residues. A number of successful synthetic strategies have been reported, including three-segment ligation,^[9d, 23] two-segment ligation^[24] and direct Fmoc-SPPS^[25]. Herein, we document its synthesis by an assembly of three peptide segments in the N→C direction, followed by a one-pot MFD protocol. Specifically, the full-length ubiquitin was disconnected into three segments at Lys27–Ala28 and Phe45–Ala46. Notably, Ala28 and Ala46 were incorporated as Cys during Fmoc-SPPS and could be desulfurized after ligation to restore the Ala side-chains (**Figure 3**). The required fragments were prepared by Fmoc-SPPS, including ubiquitin(1–27)-NHNH₂ (**16**) (**Figure S51**), ubiquitin(28–45) (**17**) containing N-terminal Cys and C-terminal hydrazide moieties (**Figure S52**), and ubiquitin(46–76) (**18**) containing N-terminal Cys (**Figure S53**). Peptide segment **16** (1.0 eq) was firstly converted into the corresponding acyl pyrazole ligated with peptide **17** (0.7 eq) in the presence of imidazole. Despite the desired ligation product **19** being the major product, C-terminal lactamization of **16** (~10%) could be observed.

After completion of the first ligation (confirmed by LC-MS), the resulting mixture was treated with additional acac at pH 3.0 to obtain intermediate **19a** with a C-terminal acyl pyrazole moiety. Segment **18** (0.63 eq) was then added and its pH was adjusted to 6.5. Upon completion, the ligation mixture was incubated with additional TCEP, *t*-BuSH and VA-044 for 2.5 h to achieve the desulfurization. Gratifyingly, after only one HPLC purification step, full-length ubiquitin (**21**) was obtained with an isolated yield of 33.6% (one-pot over five steps, average yield of 80% per step) (**Figure 3** and **Figures S54–56**). In contrast to other additive-assisted one-pot ligation-desulfurization methods, as no thiol reagents were added or generated in the reaction system our strategy did not require additional treatments such as increased temperature (40–65 °C)^[9d, 11b, 26] or prolonged reaction time (12–48 h)^[9d, 9e, 11] to obtain satisfactory desulfurization results. This case illustrates the convenience and utility of the imidazole-aided APCL strategy in facilitating sequential ligations and MFD in

one-pot, which requires no need for intermediate purification steps or other operations like extraction and heating.

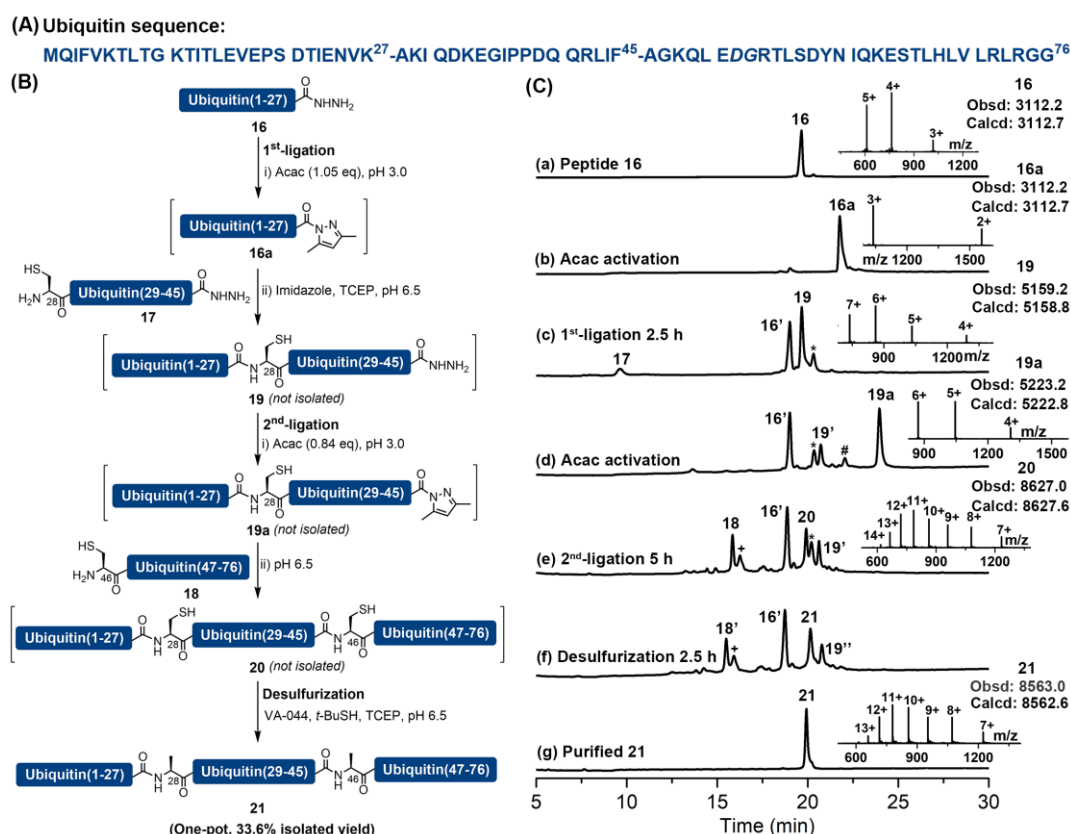


Figure 3. (A) The sequence of ubiquitin (**21**), the ligation sites are indicated with dashes, any dipeptide used in the SPPS is in italic. (B) Scheme for the synthesis of the ubiquitin **21** via a one-pot sequential ligation–MFD in the N-to-C direction. (C) HPLC traces (214 nm) and ESI–MS characterizations of the major reaction steps. (a–c) 1st ligation, **16a** is peptide acyl pyrazole derivative of **16**, **19** is the ligation product, **16'** is hydrolysis by-product and the peak * denotes C-terminal side-chain lactamization of Lys27. (d) Acac activation, **19a** denotes acylpyrazole product of **19**, **19'** is hydrolysis by-product, and the peak # denotes hydrazone intermediate. (e) 2nd ligation at 5 h, **20** is the product, the peak + corresponds to a +82 Da adduct to the remaining peptide **18**. (f) Desulfurization at 2.5 h, **21** is the desired product, **18'** and **19'** are desulfurized versions from the remaining **18** and **19'**, respectively. (g) Analytical HPLC trace and ESI-MS of the purified full-length protein **21**.

Semi-synthesis of azurin through a one-pot expressed protein ligation and refolding

The scope and utility of the APCL strategy was then extended to protein semi-synthesis, where a more challenging protein, the *Pseudomonas aeruginosa* azurin, was targeted. It is a blue copper protein, consisting of 128 amino acids, involved in electron transfer.^[27] Here, based on the available ligation site we decided that the full-length azurin could be disconnected into two segments at Phe111-Cys112. The N-terminal segment azurin(1–111)-NHNH₂ **22** (Figures S57–58) was generated through hydrazine cleavage of the corresponding *E. coli* expressed azurin(1–111)-intein fusion protein as reported literatures,^[19b, 28] while the azurin(112–128)

segment **23** containing an N-terminal Cys residue was synthesized by Fmoc-SPPS (**Figure S59**). To increase the overall synthetic yield, we aimed to carry out protein ligation and refolding in one pot. The 3-methylpyrazole enabled APCL strategy can be achieved at low pH (5.0) where Cys-containing peptide segments are more resistant to oxidation than at neutral pH, thus it offers an opportunity to perform the ligation without any reducing agent, i.e., TCEP, after which the following refolding would be allowed in one pot. To this end, the protein hydrazide segment **22** was firstly treated with acac to transform to its corresponding *N*-acyl pyrazole (**22a**) within ~1.5 h, as judged by LC-MS. Without any reducing agents like TCEP, the azurin(112–128) segment **23** and 3-methylpyrazole were added, and the pH was adjusted to 5.0 to initiate the ligation. Interestingly, the exchanged peptide *N*-acyl 3-methylpyrazole intermediate **22b** can be clearly observed with HPLC-ESI-MS, highlighting a plausible azole-exchanging and acyl-activating mechanism. As shown in **Figure 4C-d**, the ligation went to completion within 2 h, and the product **24** was co-eluted with hydrolysis and truncation (at Met13) by-products. Nevertheless, without any purification, dialysis, or ultracentrifugation treatment, a rapid dilution strategy where the ligation mixture was added directly to a redox buffer^[11b, 29] (final protein conc., ~0.3 mg/mL) was adopted, and the refolding was allowed for 24 h at 4 °C (**Figure 4C-e**). The folded protein **24a** was finally obtained by a single HPLC purification step with an isolated yield of 31.2% (over three steps). The proper folding of **24a** was confirmed by ESI-MS (~2 Da vs. **24**) (**Figure 4D, Table S5**) and fluorescence emission spectra (**Figure 4F, Figure S61**). Furthermore, the presence of a disulfide bond Cys3–Cys26 was confirmed via trypsin digestion and ESI-MS (**Figure S62**).

Lastly, the folded protein **24a** was reconstituted with Cu^{II}, exhibiting characteristic absorption band at 627 nm for the Cys112–Cu^{II} coordination (**Figure 4H, Figure S63**). Moreover, ESI-MS data (**Figure 4H, Figure S63**) also demonstrated the successful incorporation of the Cu^{II} centre, and CD spectrum (**Figure 4G, Table S6**) verified that the protein structure was maintained after the incorporation of Cu^{II}, agreeing well with the literature.^[27a, 30] Therefore, our reductant-free APCL strategy enabled by 3-methylpyrazole provides a one-pot expressed protein ligation and refolding protocol, streamlining the access of the blue-copper protein azurin.

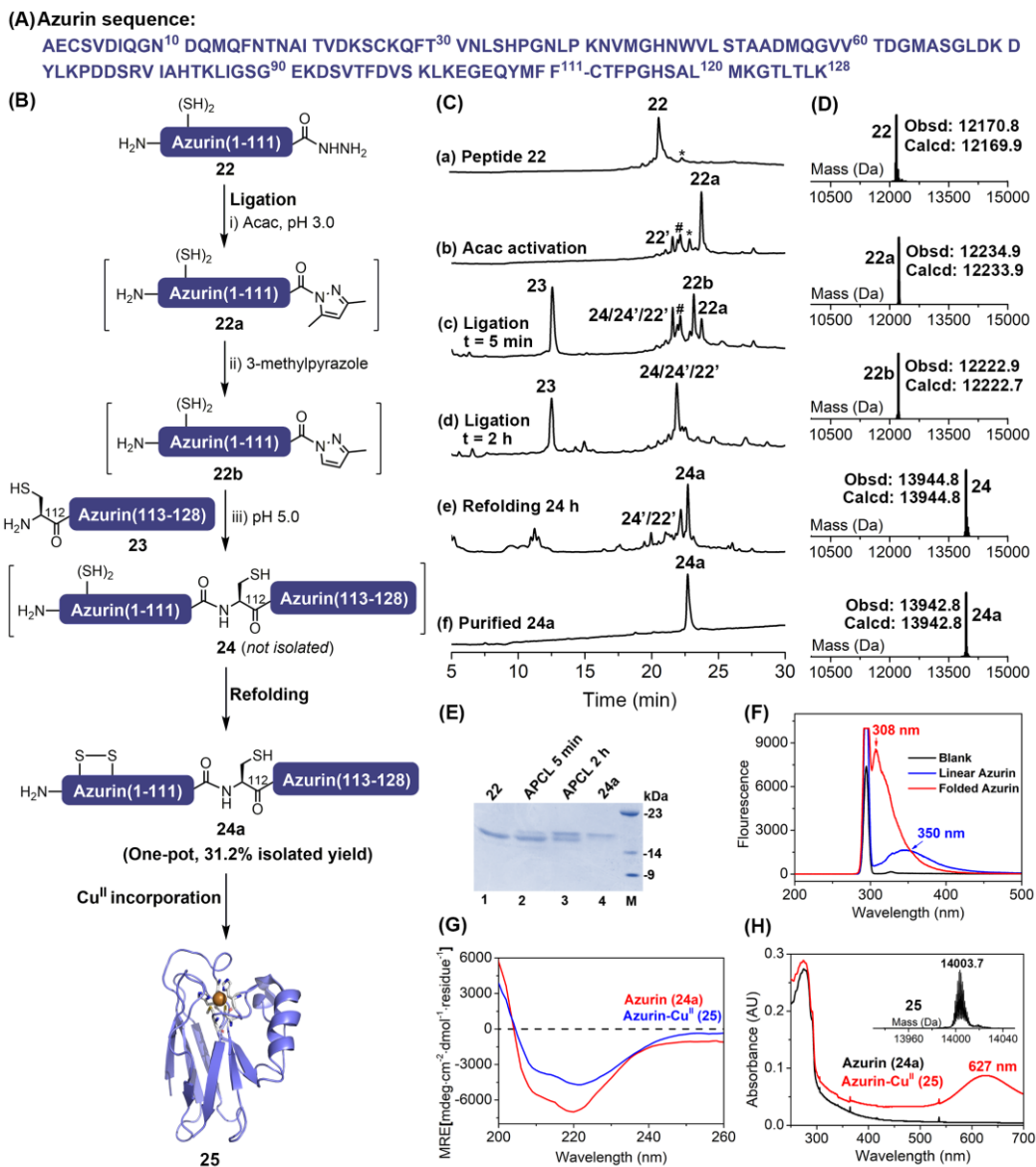


Figure 4. (A) Sequence of wild type azurin, ligation site is indicated with a dash. (B) Scheme for the semi-synthesis of the azurin **25** via one-pot expressed protein ligation-refolding with APCL. (C) HPLC traces (214 nm) of major reaction steps. (a) The purified **22**; (b) acac activation, **22a** is the acylpyrazole product, **22'** is hydrolysis by-product, the peak # denotes hydrazone intermediate and peak * denotes column contaminants; (c-d) APCL at 5 min and 2 h, **22b** is the 3-methylpyrazole substituted intermediate and **24'** is truncated versions of **24** at Met13; (e) the refolding at 24 h; (f) the purified folded protein **24a**. (D) The deconvoluted ESI-MS of the major reaction steps. (E) SDS-PAGE analysis of ligation traces and purified folded protein **24a**. (F) Fluorescence emission spectra of the linear azurin (**24**) and folded azurin (**24a**). (G) Circular dichroism spectra of the synthetic azurin (**24a**) and azurin-Cu^{II} (**25**). (H) The UV-visible absorption spectra of the apo-protein **24a** and azurin-Cu^{II} **25**, inset: the deconvoluted ESI-MS of azurin-Cu^{II} **25**.

Total synthesis of the sulfated hirudin variant 1 (sY⁶³-HV1) through the APCL strategy

Tyrosine-*O*-sulfation (sY) is a wide-spread post-translational modification, especially in salivary proteins like hirudins, the exact role of this modification has, however, been difficult to elucidate, due partially to its instability during purification from natural extract.^[31] CPS has recently evolved as an important tool in accessing sulfated proteins.^[31c, 32] Herein, we document our effort in applying the APCL strategy in the synthesis of the

hirudin variant 1 from leech bearing the labile sY modification at Tyr63, sY⁶³-HV1. For this purpose, the target protein was disconnected into two segments at Lys27–Cys28, and both were obtained by Fmoc-SPPS. The neopentyl (nP) protecting group of the sulfate was reported to be stable under typical Fmoc-SPPS and TFA cleavage conditions,^[33] hence Fmoc-Tyr(SO₃nP)-OH was chosen as the monomer. According to initial observations, the HV1(1–27) segment containing four internal Cys residues could form thiolactone almost immediately when the corresponding acyl pyrazole was activated with 3-methylpyrazole, imidazole, or MPAA^[32a]. A higher ligation pH (at 6.5) was therefore chosen and imidazole was selected as the catalyst to enable the ligation of any thiolactone, if formed. As shown in **Figure 5**, HV1(1–27) segment **26** was readily converted to its corresponding acylpyrazole form (**26a**) within 15 min, as judged by HPLC-MS. Without purification, the HV1(28–65)-sY⁶³ fragment **27**, imidazole and TCEP were added, and the pH was adjusted to 6.5, after which **26a** rapidly transformed to its thiolactone **26b**. Nevertheless, the ligation between **26b** and **27** went smoothly to afford the full-length protein **28** (**Figures S66–67**), albeit a bit slower (i.e., completed within 9 h), which agrees well with the reported method.^[32a] Interestingly, the HV1(1–27) segment did not undergo C-terminal lactamization (at Lys27) during the ligation, unlike the previous cases for peptides **9** and **16**, which suggests that this side-reaction is rather sequence specific or partially suppressed by the formation of thiolactone. It's worth mentioning that, similar to reported observations,^[32a, 33b] the nP protecting group was removed automatically during the ligation process in the presence of imidazole. In an effort to improve the overall yield, we consulted our previous work to carry out the ligation and refolding without intermediate HPLC purification.^[11b, 29] For this purpose, the ligation mixture was exchanged into a buffer containing 6 M Gdn-HCl and 0.2 M Na₂HPO₄ (pH 6.0) by ultracentrifugation (3 kDa cut-off) and was directly added dropwise into the refolding buffer (final conc. in 0.25 mg/mL). The refolding was allowed for 16 h at RT, and the refolded sY⁶³-HV1 **29** was obtained by a single HPLC purification step with an isolated yield of 39% (over three steps). The proper folding of **29** was confirmed by ESI-MS (~6 Da vs. **28**) (**Figure 5C**). For comparison, the same protein was also synthesized by conventional MPAA method (**Figures S68–69**), and more ultracentrifugation washes were needed to remove all MPAA (8 times vs 3 times for the APCL strategy), resulting in lower isolated yield (27.5%). Overall, the efficient synthesis of the sulfated sY⁶³-HV1 demonstrates that the mild APCL strategy could tolerate some of the most labile protein post-translational modifications.

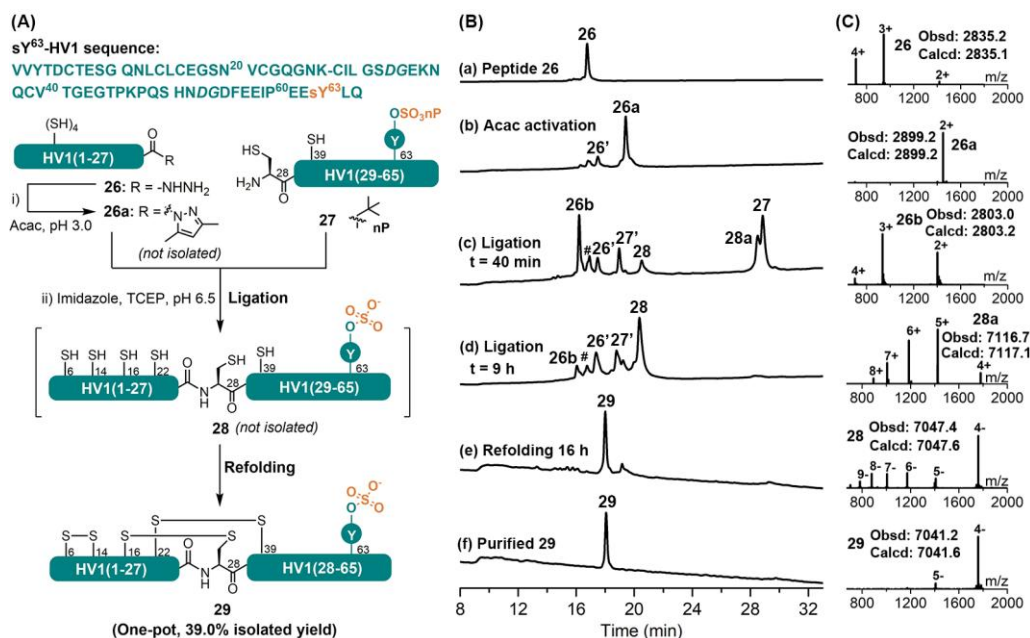


Figure 5. (A) Sequence and synthetic scheme for the sY⁶³-HV1 **29**. The ligation site is indicated with a dash and any dipeptides used are in italic. (B) HPLC traces (214 nm) of major reaction steps. (a) The purified peptide **26**; (b) acac activation, **26a** is acylpyrazole product and **26'** is hydrolysis by-product; (c-d) APCL at 40 min and 9 h, **26b** corresponds to thiolactone derivative of **26a** and **27'** corresponds to nP removed derivative of **27**. **28a** corresponds to the ligation product, **28** corresponds to nP removed product and peak # denotes unknown by-products; (e) the refolding at 16 h; (f) the purified folded product **29**. (C) ESI-MS characterizations of major reaction steps.

Synthesis of the cyclic peptide kalata B1 through the APCL strategy

Disulfide-rich cyclic peptides have high structural rigidity and are resistance against chemical, enzymatic or thermal attacks, making them promising candidates for the development of peptide-based therapeutic drugs.^[34] We were interested to explore the viability of the APCL strategy in preparing cyclic peptides. To this end, cyclization of a model peptide CGYLKFKSLA-NHNH₂ (**15**) was tested. As shown in **Figures S70–71**, both 3-methylpyrazole and imidazole gave the cyclic peptide **15c** with excellent isolated yields (76–77.4%) through APCL.

Encouraged by these positive results, we moved to synthesize a large cyclotide kalata B1 (kB1), which is from the African plant *Oldenlandia affinis*. kB1 is a head-to-tail cyclic peptide with a knotted arrangement of three disulfide bonds, which makes them extremely stable and versatile drug lead or scaffold for pharmaceutical development.^[35] Aiming to achieve a highly efficient synthetic route for kB1, the reductant-free 3-methylpyrazole-assisted APCL strategy that allow one-pot cyclization and refolding was adopted (**Figures S73–74**). In this case, the linear kB1 hydrazide **30** was treated with acac to form the corresponding acyl pyrazole fragment (**30a**) within 2 h. The cyclization was then initiated by direct addition of 3-methylpyrazole and adjustment of pH to 5.0. Again, the 3-methylpyrazole exchanged intermediate **30b** could be clearly observed, and within 3 h an extremely clean conversion to the targeted cyclization product **31** was achieved (**Figure 6B-e** and **Figures S73–74**). Without any purification and extraction steps, the resulting mixture was directly added

dropwise into the folding buffer containing 2-propanol (*i*-PrOH)^[35a] (final conc. in 0.2–0.4 mg/mL). To our delight, the oxidative folding was accomplished within 15 h, giving the folded kB1 **32** as a single peak in an excellent isolated yield (68.5% in one-pot over three steps). The proper folding of **32** was confirmed by ESI-MS (~6 Da vs. **31**) and CD spectrum (see the left inset in **Figure 6B-f**), which agrees well of reported literature.^[35a] In addition, we have synthesized kB1 with the conventional MPAA method (**Figures S75–76**), where two HPLC purification steps were required (one for the removal of MPAA and one for final purification), and the overall isolated yield (27.5%) was much lower than the current strategy. Collectively, these results demonstrate that the APCL strategy is highly viable for the synthesis of cyclic peptides, especially those bearing one or more disulfides, featuring fewer purification steps and higher synthetic yield.

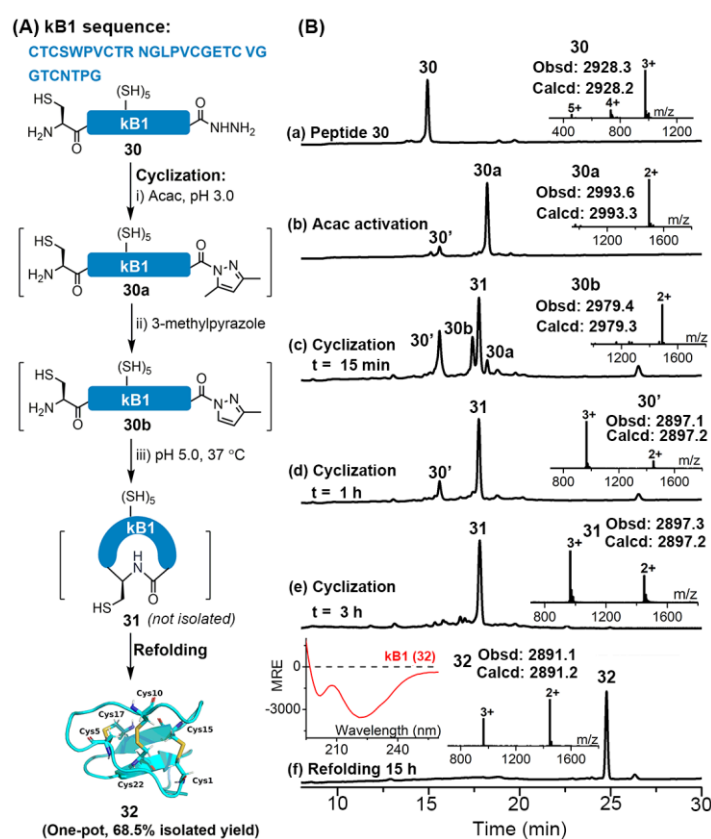


Figure 6. (A) Sequence and synthetic scheme for the cyclic peptide kB1 **32** via one-pot cyclization-refolding with APCL. (B) HPLC traces (214 nm) and ESI-MS characterizations of major reaction steps. (a) The purified **30**; (b) acac activation, **30a** is acylpyrazole and **30'** is thiolactone; (c-e) cyclization at 15 min, 1 h and 3 h, respectively. **30b** is 3-methylpyrazole substituted intermediate and **31** is cyclization product; (f) the refolding at 15 h, **32** is the folded product. Inset: CD spectrum (left) of the synthetic **32**.

Conclusion

In summary, we demonstrated that 3-methylpyrazole and imidazole can activate the weak acyl donor, peptidyl N-acylpyrazole, to enable its direct ligation with an N-terminal cysteine peptide efficiently at pH 5.0 and 6.5, respectively. Building upon the easily accessible peptidyl hydrazides (through SPPS or recombinant

expression) and its facile conversion to acylpyrazoles, the current APCL strategy is viable for both protein total and semi-synthesis. Importantly, since no peptidyl thioester or thiol additive is needed for the ligation, the APCL strategy can be conveniently combined with one-pot MFD or oxidative protein folding, which greatly improves the synthetic efficiency by avoiding the preparation of peptidyl thioester and its side-reactions (e.g., hydrolysis), reducing intermediary purification steps, and shortening overall operation time. The practicality and generality of the strategy are showcased by the total synthesis of ubiquitin via an N-to-C sequential ligation-MFD strategy and the ligation-folding of a semi-synthetic copper protein azurin, a sulfated hirudin variant, and the cyclotide kB1, all in a one-pot fashion. Given the efficiency and simplicity of APCL enabled by the commercially available and inexpensiveazole reagents, we expect it to be widely utilized to improve the efficiency of CPS, complementing the current synthetic toolbox.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos: 22277029 and 22077036). The technical assistance in fluorescence emission spectroscopy from Ms Jindi Luo was gratefully acknowledged. We also thank Ms Qi Zhang, Ms Xinyue Zhang, Mr Pengpeng Li and Mr Lan Xu for assistance with peptide synthesis.

Conflicts of interest

A Chinese patent is filed covering the strategy described in this manuscript (ZL 202310576757.4).

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords

Chemical protein synthesis • Native chemical ligation • Peptide hydrazide • Peptide *N*-acylpyrazole • Azole additive

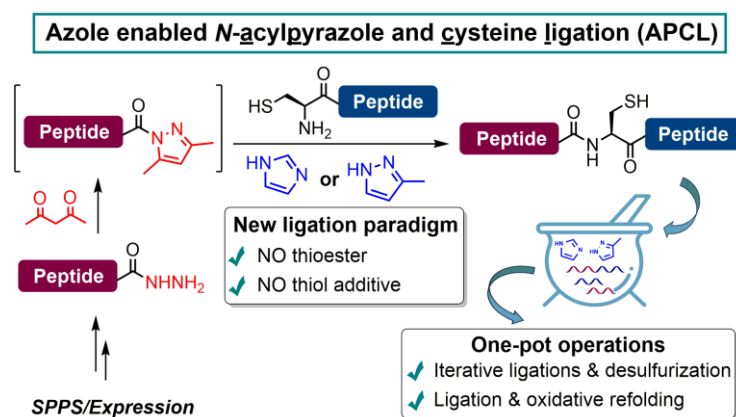
Reference

- [1] A. C. Conibear, E. E. Watson, R. J. Payne, C. F. W. Becker, *Chem. Soc. Rev.* **2018**, 47.
- [2] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, 266, 776.
- [3] a) S. B. H. Kent, *Chem. Soc. Rev.* **2009**, 38, 338; b) V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu, O. Melnyk, *Chem. Rev.* **2019**, 119.
- [4] a) S. Futaki, K. Sogawa, J. Maruyama, T. Asahara, M. Niwa, H. Hojo, *Tetrahedron Lett.* **1997**, 38; b) J. Alsina, T. S. Yokum, F. Albericio, G. Barany, *J. Org. Chem.* **1999**, 64; c) H. Hojo, Y. Kwon, Y. Kakuta, S. Tsuda, I. Tanaka, K. Hikichi, S. Aimoto, *Bull. Chem. Soc. Jpn.* **1993**, 66; d) R. von Eggelkraut-Gottanka, A. Klose, A. G. Beck-Sickinger, M. Beyermann, *Tetrahedron Lett.* **2003**, 44.
- [5] a) E. C. B. Johnson, S. B. H. Kent, *J. Am. Chem. Soc.* **2006**, 128; b) V. Diemer, O. Firstova, V. Agouridas, O. Melnyk, *Chem. Eur. J.* **2022**, 28.

- [6] a) Q. Wan, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2007**, *46*, 9248; b) C. Haase, H. Rohde, O. Seitz, *Angew. Chem., Int. Ed.* **2008**, *47*; c) J. Chen, Q. Wan, Y. Yuan, J. Zhu, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2008**, *47*; d) K. S. Ajish Kumar, M. Haj-Yahya, D. Olschewski, H. A. Lashuel, A. Brik, *Angew. Chem., Int. Ed.* **2009**, *48*; e) L. R. Malins, R. J. Payne, *Aust. J. Chem.* **2015**, *68*; f) H. M. Burke, L. McSweeney, E. M. Scanlan, *Nat. Commun.* **2017**, *8*; g) Y. Gui, L. Qiu, Y. Li, H. Li, S. Dong, *J. Am. Chem. Soc.* **2016**, *138*.
- [7] H. Rohde, J. Schmalisch, Z. Harpaz, F. Diezmann, O. Seitz, *ChemBioChem* **2011**, *12*.
- [8] T. Moyal, H. P. Hemantha, P. Siman, M. Refua, A. Brik, *Chem. Sci.* **2013**, *4*.
- [9] a) P. Siman, O. Blatt, T. Moyal, T. Danieli, M. Lebediker, H. A. Lashuel, A. Friedler, A. Brik, *ChemBioChem* **2011**, *12*; b) R. E. Thompson, X. Liu, N. Alonso-García, P. J. B. Pereira, K. A. Jolliffe, R. J. Payne, *J. Am. Chem. Soc.* **2014**, *136*; c) M. Haj-Yahya, H. A. Lashuel, *J. Am. Chem. Soc.* **2018**, *140*; d) Y.-C. Huang, C.-C. Chen, S. Gao, Y.-H. Wang, H. Xiao, F. Wang, C.-L. Tian, Y.-M. Li, *Chem. Eur. J.* **2016**, *22*; e) P. W. Erickson, J. M. Fulcher, P. Spaltenstein, M. S. Kay, *Bioconjug. Chem.* **2021**, *32*; f) K. Ohkawachi, D. Kobayashi, K. Morimoto, A. Shigenaga, M. Denda, K. Yamatsugu, M. Kanai, A. Otaka, *Org. Lett.* **2020**, *22*; g) S. Suzuki, Y. Nakajima, N. Kamo, A. Osakabe, A. Okamoto, G. Hayashi, H. Murakami, *Molecules* **2023**, *28*.
- [10] a) D. Bang, B. L. Pentelute, S. B. H. Kent, *Angew. Chem. Int. Ed.* **2006**, *45*; b) T. Durek, V. Y. Torbeev, S. B. H. Kent, *Proc. Natl. Acad. Sci.* **2007**, *104*; c) E. E. Watson, X. Liu, R. E. Thompson, J. Ripoll-Rozada, M. Wu, I. Alwis, A. Gori, C.-T. Loh, B. L. Parker, G. Otting, S. Jackson, P. J. B. Pereira, R. J. Payne, *ACS Cent. Sci.* **2018**, *4*.
- [11] a) K. Sakamoto, S. Tsuda, M. Mochizuki, Y. Nohara, H. Nishio, T. Yoshiya, *Chem. Eur. J.* **2016**, *22*; b) P. Liao, C. He, *Front. Chem.* **2021**, *9*, 735149.
- [12] T. S. Chisholm, D. Clayton, L. J. Dowman, J. Sayers, R. J. Payne, *J. Am. Chem. Soc.* **2018**, *140*.
- [13] a) K. Kitagawa, H. Adachi, Y. Sekigawa, T. Yagami, S. Futaki, Y. J. Gu, K. Inoue, *Tetrahedron* **2004**, *60*; b) J. Alsina, S. A. Kates, G. Bárány, F. J. M. i. m. b. Albericio, *Methods Mol. Biol.* **2005**, *298*; c) J. B. Blanco-Canosa, P. E. Dawson, *Angew. Chem., Int. Ed.* **2008**, *47*; d) J. A. Camarero, B. J. Hackel, J. J. de Yoreo, A. R. Mitchell, *J. Org. Chem.* **2004**, *69*; e) M. Yanase, K. Nakatsu, C. J. Cardos, Y. Konda, G. Hayashi, A. Okamoto, *Chem. Sci.* **2019**, *10*; f) J. Dheur, N. Ollivier, O. Melnyk, *Org. Lett.* **2011**, *13*; g) H. E. Elashal, Y. E. Sim, M. Raj, *Chem. Sci.* **2017**, *8*; h) B. Cowper, T. M. Sze, B. Premdjee, A. F. Bongat White, A. Hacking, D. Macmillan, *Chem. Commun.* **2015**, *51*; i) B. Yan, W. Shi, L. Ye, L. Liu, *Curr. Opin. Chem. Biol.* **2018**, *46*; j) G.-M. Fang, Y.-M. Li, F. Shen, Y.-C. Huang, J.-B. Li, Y. Lin, H.-K. Cui, L. Liu, *Angew. Chem., Int. Ed.* **2011**, *50*.
- [14] K. Sakamoto, S. Tsuda, H. Nishio, T. Yoshiya, *Chem. Commun.* **2017**, *53*.
- [15] a) Y. Wang, L. Han, N. Yuan, H. Wang, H. Li, J. Liu, H. Chen, Q. Zhang, S. Dong, *Chem. Sci.* **2018**, *9*; b) Y. Dao, B. Wang, W. Dong, J. Zhang, C. Zhong, Z. Zhang, S. Dong, *Chin. J. Chem.* **2021**, *39*.
- [16] J.-S. Zheng, S. Tang, Y.-K. Qi, Z.-P. Wang, L. Liu, *Nat. Protoc.* **2013**, *8*.
- [17] a) J. Thom, D. Anderson, J. McGregor, G. Cotton, *Bioconjug. Chem.* **2011**, *22*; b) D. Zhang, Z. Wang, S. Hu, N.-Y. Chan, H. T. Liew, J. Lescar, J. P. Tam, C.-F. Liu, *Bioconjug. Chem.* **2022**, *33*; c) A. L. Adams, B. Cowper, R. E. Morgan, B. Premdjee, S. Caddick, D. Macmillan, *Angew. Chem., Int. Ed.* **2013**, *52*; d) Y. Kajihara, Y. Kanemitsu, M. Nishihara, R. Okamoto, M. Izumi, *J. Pept. Sci.* **2014**, *20*; e) Y. Qiao, G. Yu, K. C. Kratch, X. A. Wang, W. W. Wang, S. Z. Leeuwon, S. Xu, J. S. Morse, W. R. Liu, *J. Am. Chem. Soc.* **2020**, *142*; f) L. Shi, H. Chen, S.-Y. Zhang, T.-T. Chu, Y.-F. Zhao, Y.-X. Chen, Y.-M. Li, *J. Pept. Sci.* **2017**, *23*; g) Z. Mo, S. Lin, W. Chen, C. He, *Angew. Chem., Int. Ed.* **2022**, *61*.
- [18] D. T. Flood, J. C. J. Hintzen, M. J. Bird, P. A. Cistrone, J. S. Chen, P. E. Dawson, *Angew. Chem., Int. Ed.* **2018**, *57*.
- [19] a) Y. Li, J. Liu, Q. Zhou, J. Zhao, P. Wang, *Chin. J. Chem.* **2021**, *39*; b) S. S. Kulkarni, E. E. Watson, J. W. C. Maxwell, G. Niederacher, J. Johansen-Leete, S. Huhmann, S. Mukherjee, A. R. Norman, J. Kriegesmann, C. F. W. Becker, R. J. Payne, *Angew. Chem., Int. Ed.* **2022**, *61*.
- [20] a) H. A. Staab, A. Mannschreck, *Chem. Ber.* **1962**, *95*; b) A. R. Katritzky, S. Hoffmann, K. Suzuki, *ARKIVOC* **2004**; c) T. Mino, S. Sakamoto, I. Hamachi, *Biosci. Biotechnol. Biochem.* **2021**, *85*.
- [21] Y. Li, A. Yongye, M. Giulianotti, K. Martinez-Mayorga, Y. Yu, R. A. Houghten, *J. Comb. Chem.* **2009**, *11*.
- [22] a) P. Siman, S. V. Karthikeyan, M. Nikolov, W. Fischle, A. Brik, *Angew. Chem., Int. Ed.* **2013**, *52*; b) F.-K. Deng, L. Zhang, Y.-T.

- Wang, O. Schneewind, S. B. H. Kent, *Angew. Chem., Int. Ed.* **2014**, 53; c) H. Sun, A. Brik, *Acc. Chem. Res.* **2019**, 52.
- [23] D. Bang, G. I. Makhatadze, V. Tereshko, A. A. Kossiakov, S. B. Kent, *Angew. Chem., Int. Ed.* **2005**, 44.
- [24] a) K. S. A. Kumar, L. Spasser, L. A. Erlich, S. N. Bavikar, A. Brik, *Angew. Chem., Int. Ed.* **2010**, 49; b) L. A. Erlich, K. S. A. Kumar, M. Haj-Yahya, P. E. Dawson, A. Brik, *Org. Biomol. Chem.* **2010**, 8.
- [25] a) R. Ramage, J. Green, T. W. Muir, O. M. Ogunjobi, S. Love, K. Shaw, *Biochem. J.* **1994**, 299; b) F. El Oualid, R. Merkx, R. Ekkebus, D. S. Hameed, J. J. Smit, A. de Jong, H. Hilkmann, T. K. Sixma, H. Ovaa, *Angew. Chem., Int. Ed.* **2010**, 49; c) S. N. Bavikar, L. Spasser, M. Haj-Yahya, S. V. Karthikeyan, T. Moyal, K. S. Ajish Kumar, A. Brik, *Angew. Chem., Int. Ed.* **2012**, 51.
- [26] J. Schmalisch, O. Seitz, *Chem. Commun.* **2015**, 51.
- [27] a) D. K. Garner, M. D. Vaughan, H. J. Hwang, M. G. Savelieff, S. M. Berry, J. F. Honek, Y. Lu, *J. Am. Chem. Soc.* **2006**, 128; b) T. D. Wilson, Y. Yu, Y. Lu, *Coord. Chem. Rev.* **2013**, 257.
- [28] J. Liu, O. Ekanayake, D. Santoleri, K. Walker, S. Rozovsky, *ChemBioChem* **2020**, 21.
- [29] P. Liao, H. Liu, C. He, *Chem. Sci.* **2022**, 13.
- [30] a) Y. Yu, N. M. Marshall, D. K. Garner, M. J. Nilges, Y. Lu, *J. Inorg. Biochem.* **2022**, 234; b) N. Bonander, B. Göran Karlsson, T. Vänngård, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1995**, 1251.
- [31] a) W. B. Huttner, *Nature* **1982**, 299; b) M. J. Stone, R. J. Payne, *Acc. Chem. Res.* **2015**, 48; c) J. W. C. Maxwell, R. J. Payne, *Curr. Opin. Chem. Biol.* **2020**, 58.
- [32] a) Y. Yang, M. Liang, R. Wang, C. He, *Chin. Chem. Lett.* **2023**, 34; b) D. Clayton, S. S. Kulkarni, J. Sayers, L. J. Dowman, J. Ripoll-Rozada, P. J. B. Pereira, R. J. Payne, *RSC Chem. Biol.* **2020**, 1; c) R. E. Thompson, X. Liu, J. Ripoll-Rozada, N. Alonso-García, B. L. Parker, P. J. B. Pereira, R. J. Payne, *Nat. Chem.* **2017**, 9; d) E. E. Watson, J. Ripoll-Rozada, A. C. Lee, M. C. L. Wu, C. Franck, T. Pasch, B. Premdjee, J. Sayers, M. F. Pinto, P. M. Martins, S. P. Jackson, P. J. B. Pereira, R. J. Payne, *Proc. Natl. Acad. Sci.* **2019**, 116.
- [33] a) L. S. Simpson, J. Z. Zhu, T. S. Widlanski, M. J. Stone, *Chem. Biol.* **2009**, 16; b) C. Li, C. He, *Org. Biomol. Chem.* **2020**, 18.
- [34] a) M. L. Colgrave, D. J. Craik, *Biochemistry* **2004**, 43; b) D. J. Craik, J. S. Mylne, N. L. Daly, *Cell. Mol. Life Sci.* **2010**, 67.
- [35] a) N. L. Daly, S. Love, P. F. Alewood, D. J. Craik, *Biochemistry* **1999**, 38; b) J.-S. Zheng, S. Tang, Y. Guo, H.-N. Chang, L. Liu, *ChemBioChem* **2012**, 13; c) X. Jia, S. Kwon, C.-I. A. Wang, Y.-H. Huang, L. Y. Chan, C. C. Tan, K. J. Rosengren, J. P. Mulvenna, C. I. Schroeder, D. J. Craik, *J. Biol. Chem.* **2014**, 289; d) S. J. de Veer, M.-W. Kan, D. J. Craik, *Chem. Rev.* **2019**, 119.

Entry for the Table of Contents



A novel strategy is herein introduced which utilizes azole additives like 3-methylpyrazole and imidazole to activate the weak acyl donor peptidyl *N*-acylpyrazole, and thereby enables its direct ligation to an N-terminal cysteine peptide. It eliminates the requirement for peptidyl thioester and thiol additives in the existing paradigm and allows for the following desulfurization or refolding in a one-pot protocol.