Article

Aldehyde Capture Ligation for Synthesis of Native Peptide Bonds

Monika Raj, Huabin Wu, Sarah L. Blosser, Marc A. Vittoria, and Paramjit S. Arora

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.5b03538 • Publication Date (Web): 12 May 2015

Aldehyde Capture Ligation for Synthesis of Native Peptide Bonds

Monika Raj,[‡,†] Huabin Wu,[‡] Sarah L. Blosser, Marc A. Vittoria, and Paramjit S. Arora*

Department of Chemistry, New York University, 100 Washington Square East, New York, NY 10003

Supporting Information Placeholder

ABSTRACT: Chemoselective reactions for amide bond formation have transformed the ability to access synthetic proteins and other bioconjugates through ligation of fragments. In these ligations, amide bond formation is accelerated by transient enforcement of an intramolecular reaction between the carboxyl and the amine termini of two fragments. Building on this principle, we introduce an aldehyde capture ligation that parlays the high chemoselective reactivity of aldehydes and amines to enforce amide bond formation between amino acid residues and peptides that are difficult to ligate by existing technologies.

INTRODUCTION

Methods for chemoselective formation of amide bonds between unprotected peptide fragments have provided valuable access to recombinant proteins and have become reliable methods for the synthesis of bioconjugates.[1-9] These advances have made total synthesis of therapeutic peptides and proteins, hormones, and modified antibodies a credible objective.[10-17] The premise underlying peptide ligation approaches is that the amide bond formation step can be accelerated by capturing the carboxyl and the amine functionalities and enforcing intramolecular bond formation. Various ligation technologies[18-24] including native chemical ligation (NCL)[25] and the Staudinger ligation,[26,27] as well as auxiliary based methods,[5,28,29] utilize this concept and are widely used. Here we describe an approach termed the aldehyde capture ligation, or ACL, that exploits a similar fundamental logic to the auxiliary based methods first conceived three decades ago by Kemp and others.[28-34]

Kemp and Vellaccio postulated that an amine nucleophile may be captured by an aldehyde group in proximity of a carboxylic acid masked as an ester to drive its aminolysis.[32] Reactions of 8-acetoxy-1-naphthaldehyde or 2-acetoxybenzaldehyde with benzylamine provided a mixture of the desired amide as well as imine byproducts in a solvent dependent manner.[29] We postulated that the amount of the dehydrated imine byproduct could be reduced if the carboxylic acid was masked as a more reactive thio or seleno ester.[34] If successful, the method will, in principle, be applicable to any N-terminal amino acid residue for amide bond formation. For example, NCL requires the presence of a cysteine group as the N-terminal residue but ACL should work for any amine because the capture step is changed from thioester exchange, as in NCL, to aldehyde–amine condensation (Figure 1a).

Aldehyde capture ligation (ACL),[28,30-32] envisions an *o*-benzaldehyde ester to enforce an intramolecular reaction between the carboxyl and the amine partners (Figure 1b).[29] The *o*-benzaldehyde ester (**I**) could condense with the amine functionality to form the hemiaminal (**II**), hemiaminal ester (**III**) or imine (**IV**) intermediate.[32] Subsequent rearrangement of the hemiaminal/imine/hemiaminal ester intermediate through a cyclic transition state would yield the amide product (**V**) and the *ortho*-benzaldehyde auxiliary (**VI**). The *ortho*-aldehyde group in the auxiliary is designed to reversibly capture the amine and intramolecularly deliver it to an activated acid. The esters may also directly condense with the amine;[35-38] although our studies suggest that this is not a major pathway. The mechanistic intermediates have been postulated by Kemp[33] (Figure 2a) and also evaluated by Ito.[34] As the earlier studies by Kemp suggest that the imine is a nonproductive intermediate, we postulated that dehydration of the initially formed hemiaminal **II** would need to be slower than the X→N shift to yield the desired amide in high yield. If the hemiaminal is indeed the functional intermediate rather than an imine, we hypothesized that secondary amines may also undergo aldehyde capture ligation.

Figure 2 compares ACL with three other aldehyde capture approaches, including Kemp's, that enforce amide bond formation.[29] Tam and coworkers utilized α-acyloxyacetaldehyde auxiliary to activate C-terminal carboxylic acid groups (Figure 2b). Condensation of the activated acid with peptides bearing *N*-terminal cysteine, serine or threonine residues yield psuedoproline-linked fragments.[8,39,40] Li et al. have recently demonstrated that *O*-salicylaldehyde ester enables coupling of N-terminal serine/threonine fragments (Figure 2c).[30,31] Serine-threonine ligation and ACL (Figure 2d) utilize a similar benzaldehyde auxiliary; however, we envisioned broader substrate scope for the ACL if the initially formed hemiaminal could be acylated directly.

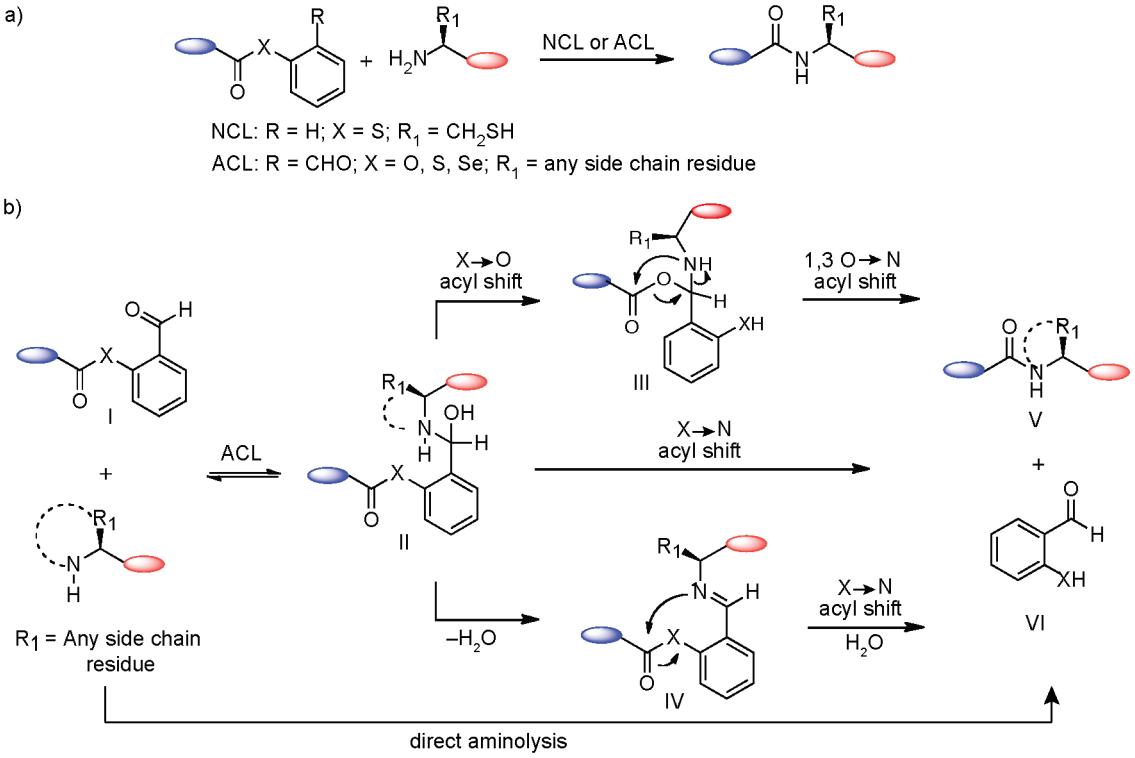


Figure 1. Rationale for the aldehyde capture ligation (ACL). (a) Comparison of a phenylthioester-initiated native chemical ligation (NCL) and ACL. ACL seeks to accelerate amide bond formation through transient condensation of the amine partner with an appended aldehyde group and enforcement of a cyclic intermediate. NCL enhances the effective molarity between reacting partners through a transient thioester bond formation with the cysteine side chain. (b) Several discreet mechanisms may be postulated for ACL.

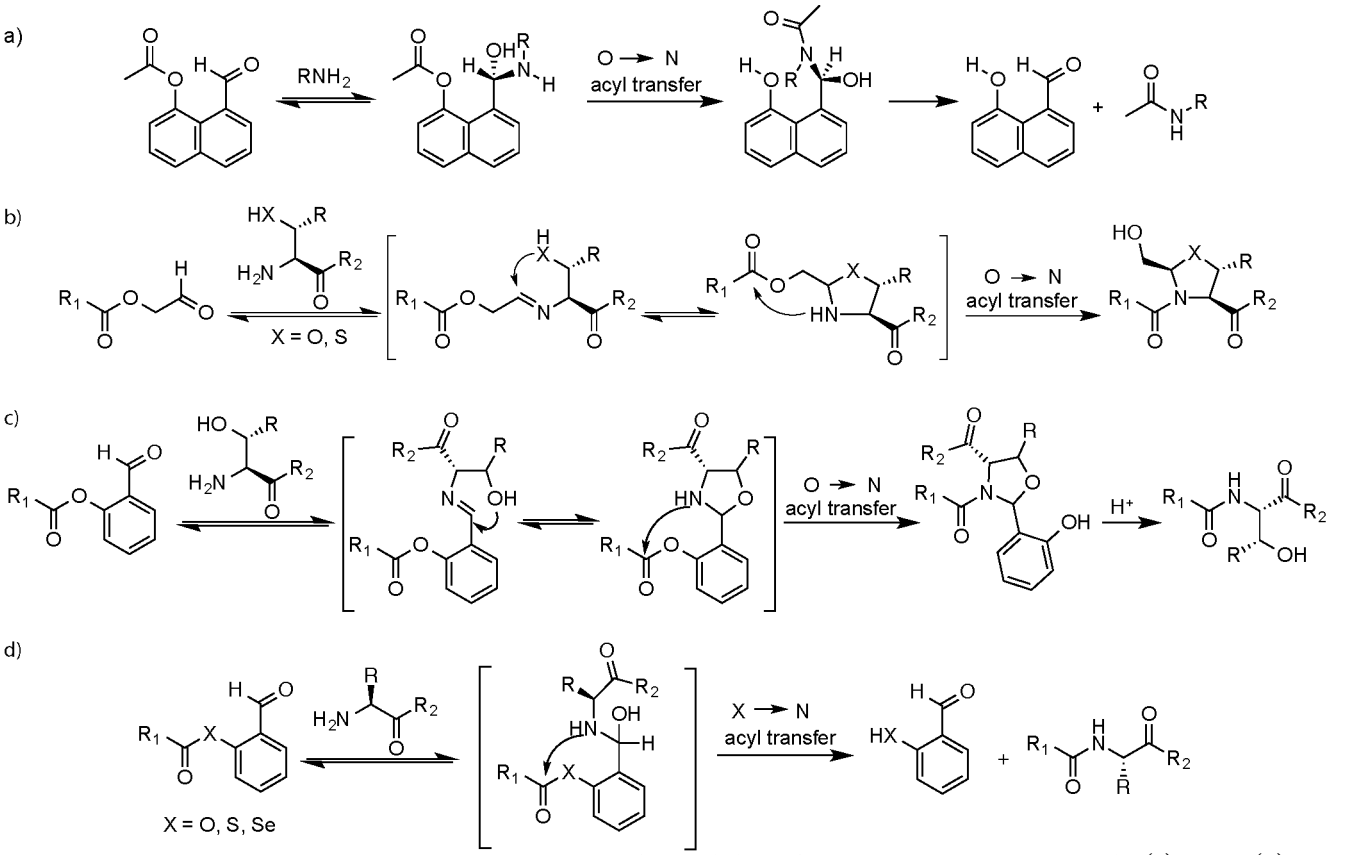


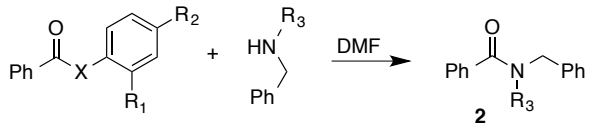
Figure 2. Mechanistic approaches that utilize aldehyde capture for amide bond formation. Approaches outlined by (a) Kemp, (b) Tam, and (c) Li are compared with (d) ACL.

RESULTS AND DISCUSSION

Evaluation of the reactivity of oxo-, thio-, and seleno-benzaldehydesters for amide bond formation

We began the ACL auxiliary development investigations by condensing salicylaldehyde-derived benzylester **1a** with benzylamine in DMF (Table 1).[28,32] Complete conversion of the ester to amide **2** was observed in less than five hours as monitored by LCMS. Presence of the *ortho*-aldehyde group is critical for the ACL mechanism as phenyl and *para*-benzaldehyde esters, **1b** and **1c**, respectively, react at much slower rates. Compound **1c** also serves as a control for evaluating potential reactivity increases due to the presence of an electron-withdrawing group at the *ortho* or *para* position of the aromatic ring. Aminolysis reactions are not, considered to be sensitive to minor pK_b differences in the leaving phenolate ions, as the formation of the tetrahedral transition state is the rate limiting step; although, analysis is complicated by proton transfers.[41-44] The pKa's of the *ortho*-formylphenol is slightly lower than the *para* substituted analog while both pKa's are much lower than that of the basic amine nucleophile. The *para*-benzaldehyde ester analog, **1c**, undergoes amide bond formation much slower than the *ortho* analog **1a** reflecting the faster rate of aldehyde-amine union than direct acylation of the substituted phenylesters. While the transient hemiaminal/imine in the *ortho* analog is rapidly captured for acylation, a similar step is not possible for the *para* analog.

Table 1. Model studies for the development of ACL



2

Ester	X	R^1	R^2	R^3	Time[c]
1a[a]	O	CHO	H	H	5 h
1b[a]	O	H	H	H	>40 h
1c[a]	O	H	CHO	H	>40 h
1d[a]	S	CHO	H	H	10 min
1e[a]	S	H	H	H	60 min
1f[a]	S	H	CHO	H	10 h
1g[a]	Se	CHO	H	H	<2 min[d]
1g[b]	Se	CHO	H	H	<2 min[d]
1h[b]	Se	H	CHO	H	>60 h
1i[b]	Se	H	H	H	30 min
1g[b]	Se	CHO	H	CH_3	<2 min[d]
1h[b]	Se	H	CHO	CH_3	>60 h
1i[b]	Se	H	H	CH_3	30 min
1j[b]	Se	NO_2	H	H	<2 min[d]

[a]Reaction conditions: esters (50 μmol) and amines (125 μmol) in 1mL DMF. [b]Reaction conditions: esters (2 μmol) and amines (5 μmol) in 1mL DMF. [c]Time for >95 % conversion of starting ester to the amide product at room temperature as determined by LCMS. [d]These rapid reaction rates could not be accurately measured with the HPLC assay.

Next, we sought to enhance the efficiency of the ACL reaction by converting the leaving group from a phenol moiety to a thiophenol group. Importance of *ortho*-aldehyde group for ACL is also evident in the thioester series **1d-1f**, as **1d**, with an *ortho*-aldehyde group, reacts more efficiently than the unsubstituted thiophenyl and *para*-thiobenzaldehyde esters, **1e** and **1f**, respectively (Table 1). Extensive analysis of aminolysis of thioesters suggests that *N*-acylations also proceed through tetrahedral intermediates as oxoesters.[44,45] (The thiobenzaldehyde ester derivatives undergo a side reaction; this reaction can be suppressed by converting the benzaldehyde to the corresponding benzophenone. Details on this benzophenone-analog are included in the Supporting Information, Figure S1).

To further accelerate the ACL reaction, we changed the leaving group from a thiophenol moiety to a selenophenol derivative **1g** (Table 1).[35,37,38] As expected, substitution with selenol provided a additional boost to the reaction rate with complete conversion to amide **2** occurring faster than could be measured under the reaction conditions. The above reactions were performed using 50 mM of the oxo/thio/selenoesters and 2.5 eq of benzylamine. To evaluate the difference in reactivity of different selenoester derivatives, we utilized a lower concentration of selenoester (2 mM). A 2.5 eq excess of benzylamine was uniformly maintained in all experiments (Supporting Information, Figure S2). (Under our HPLC conditions designed for qualitative assessment of reaction rates, we could not accurately measure the rates of reactions that completed in less than 2 minutes; further lowering of substrate concentrations to reduce reaction rates also affects accurate detection of starting materials and products.) Under these conditions, we find selenoester **1g** with the *o*-benzaldehyde group undergoes complete conversion to the amide in less than 2 min, with rate substantially faster than the *para* and unsubstituted benzylselenoester analogs **1h** (> 60 h) and **1i** (30 min), respectively. *The exceptionally slow rate of amide formation with the para-selenoester analog **1h** suggests that condensation of aldehydes with amines is faster than aminolysis of selenoesters.* We postulate that the *para*-aldehyde group scavenges the amine and lowers the concentration of the free amine impacting the rate of amide bond formation; however, because we used an excess of the amine, the amide bond should have formed at a faster rate if the rate of direct acylation (with aldehyde substituted benzyl esters) were competitive with aldehyde/amine condensation.

We also investigated if secondary amines are able to utilize the ACL mechanism. A similar trend in reaction rates was observed with the *N*-methylbenzylamine as above: the *ortho*-aldehyde analog **1g** undergoes selenoester aminolysis much faster than the *para*-aldehyde (**1h**) or the unsubstituted selenoester (**1i**). Reactivity of secondary amines result suggests that hemiaminal intermediate is engaging the ester, as the imine formed from a secondary amine would be expected to be less nucleophilic.

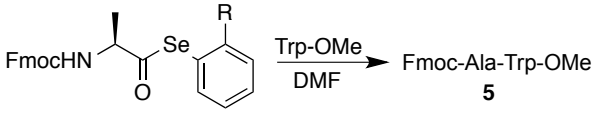
Lastly, we tested the potential of a different electron withdrawing group in place of the aldehyde group to mediate coupling of selenoesters with benzylamine through direct aminolysis.[34,46] Compound **1j**, which features an *ortho*-nitrophenyl selenoester, undergoes aminolysis on par with *ortho*-benzaldehyde selenoester; although accurate comparisons are not possible under our HPLC conditions. The pKa of *ortho*-nitrophenol (7.23) falls between that of *ortho*-formylphenol (6.79) and *para*-formylphenol (7.66) thus the rate of aminolysis for the nitro derivative may not be readily explained by the leaving group ability of the particular selenols, assuming the pKa trend from phenols translates to selenol derivatives. This observation is consistent

with literature reports that the formation and not the breakdown of the tetrahedral transition state is the rate limiting step in ester aminolysis reactions, as discussed above.[41-44] Further studies, *vide infra*, with the benzaldehyde selenoesters and nitrophenyl selenoesters indicate that subtle electronic effects may be modulating the reactivity of the benzaldehyde selenoesters. Overall, although, we cannot rule out direct acylation of amines as a contributor to amide bond formation, our model studies support the aldehyde capture mechanism for *ortho*-selenobenzaldehyde derivatives with a hemiaminal as the major pathway.

Application of ACL to amino acid derivatives: Evaluation of potential epimerization

We next sought to determine if amino acid derived selenobenzaldehyde esters can provide the desired dipeptides in a chemoselective and efficient manner. The ACL reaction between alanine-derived selenobenzaldehyde ester, **3**, and tryptophan methyl ester follows a second order rate constant, (FmocAla-COSe-*o*-PhCHO, $k = 1.65 \pm 0.06\ M^{-1}S^{-1}$) (Table 2 and Supporting Information, Figure S3). This reaction undergoes complete conversion to the dipeptide roughly 100-times faster as compared to the selenoester analog without the *o*-aldehyde group (FmocAla-COSePh, $k = 0.015 \pm 0.001\ M^{-1}S^{-1}$).

Table 2. Kinetics of the aldehyde capture ligation for dipeptide formation[a]



Selenoester	R	k [$M^{-1}S^{-1}$]	Relative Rates
3	CHO	1.653 ± 0.0587	110
4	H	0.015 ± 0.0011	1

[a]Conditions: FmocAla-seleno-ester, 3 or 4 (2 μmol), HCl·NH_2-Trp-OMe (2 μmol) and Et_3N (2 μmol) in 1mL DMF.

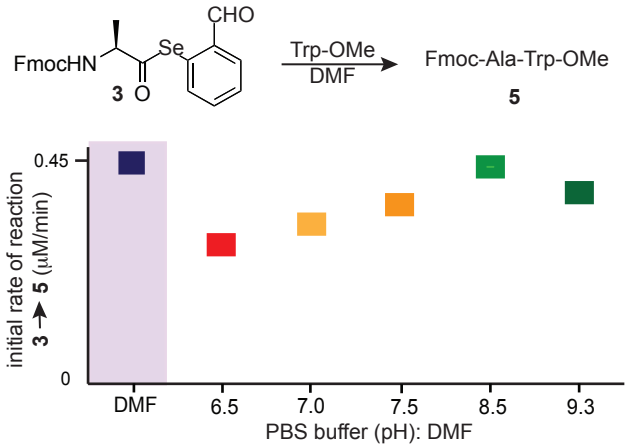


Figure 3. Effect of pH on reaction rates. Conditions: FmocAla-COSe-*o*-PhCHO, **3** (2 μmol), HCl·NH_2-Trp-OMe (4 μmol) and Et_3N (4 μmol) in 1mL DMF or 1:1, DMF:100 mM NaH_2PO_4 buffer (pH ~ 6.5-9.3). Plotted values are the average of two independent experiments.

We determined the compatibility of ACL for synthesis of peptides in aqueous buffers (with DMF as a co-solvent) for potential applications aimed at synthesis of large peptides and proteins or other bioconjugates by fragment coupling. The formation and reactions of hemiaminals and imines are known to be sensitive to buffer pH;[47] however, hydrolysis of selenoesters is also pH dependent. Our results show that initial rates of dipeptide formation were higher at basic pH, with pH 8.5 proving to be optimal (Figure 3 and Figure S4).

In these exploratory studies designed to determine the scope of the aldehyde capture ligation, we initially prepared amino acid and peptide thio- and seleno-benzaldehyde esters by coupling activated carboxylic acids and *o*-benzaldehyde disulfide (**6**) or diselenide (**7**) in the presence of a reducing agent, such as TCEP (Figure 4). The diselenide, itself, is available in one step from 2-bromobenzaldehyde (Supplementary Information).[48] (We have also explored solid phase synthesis of peptide selenoesters as described below.)

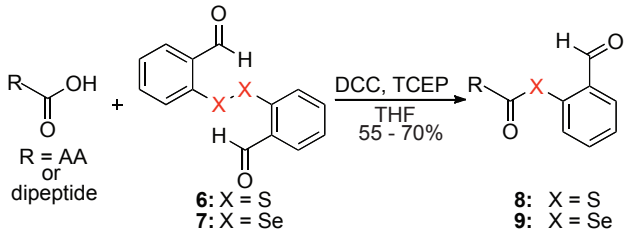


Figure 4. Synthesis of seleno- and thio-*o*-benzaldehyde esters from Fmoc amino acids (AA) or peptides.

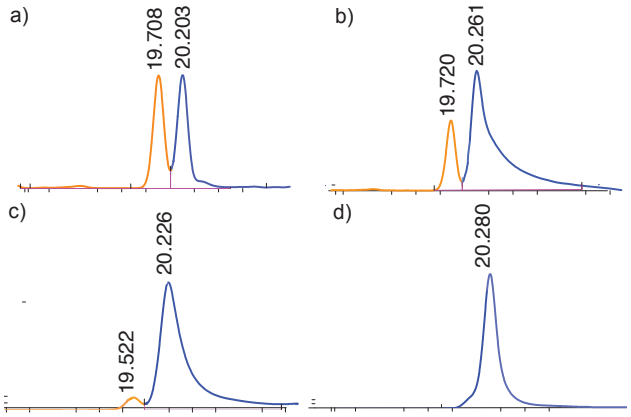


Figure 5. HPLC studies to determine potential epimerization of the amino acid selenoester: (a) Analytical HPLC trace of a 1:1 mixture of diastereoisomers, FmocV-*D*A-AQ and Fmoc-V-*L*A-AQ. (b) Analytical HPLC trace of a 1:4 mixture of diastereoisomers, FmocV-*D*A-AQ and Fmoc-V-*L*A-AQ. (c) Analytical HPLC trace of a 1:49 mixture of diastereoisomers, FmocV-*D*A-AQ and FmocV-*L*A-AQ (d) Crude HPLC trace of ACL reaction mixture FmocVal-*L*Ala-seleno-*ortho*-benzaldehyde ester (10 μmol), TFA·Ala-Gln-$CONH_2$ (20 μmol) and Et_3N (20 μmol) in 1 mL DMF. HPLC Conditions: 0.1 % TFA (v/v) in water (solvent A): acetonitrile (solvent B); gradient 35-65% in 60 min, flow rate = 0.5 mL/min.

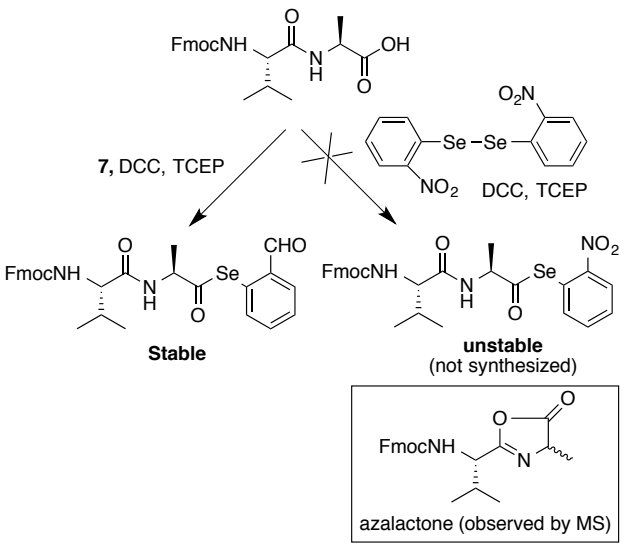
Epimerization of activated amino acids is a critical concern in peptide synthesis. The selenobenzaldehyde-modified peptide

may be expected to undergo epimerization during the course of aldehyde capture ligation possibly through the formation of an azalactone. To investigate this possibility and quantify the amount of epimerized product, we prepared a tetrapeptide FmocVal-*L*Ala-Ala-Gln-NH_2 and the expected epimerized analog, FmocVal-*D*Ala-Ala-Gln-NH_2. These products would result from ligation of FmocVal-*L*Ala-COSe-*o*-PhCHO with NH_2-Ala-Gln-$CONH_2$. An HPLC assay that allows detection of as low as 2% epimerization was developed; careful analysis showed the absence of racemization under the reaction conditions (Figure 5 and Figure S5).

Condensation of FmocVal-Ala selenobenzaldehyde with NH_2-Ala-Gln-$CONH_2$ is a relatively fast reaction. To rule out the possibility that the epimerization is occurring but at a slower rate, we tested condensation of FmocAla-Val-*L*-COSe-*o*-PhCHO with NH_2-Ala-Gln-$CONH_2$ under the same reaction conditions. Amide bond formation involving valine selenobenzaldehyde esters is relatively slow (Table 3, entries 8 and 9). Roughly 2% epimerization was observed for this slower ACL coupling (Figure S5).

The low level of epimerization observed for the ACL reactions was surprising because activated acids, especially in the context of peptides where the azalactone formations are often observed as an intermediate to epimerization.[49] Above, we described aminolysis experiments with the nitrophenyl selenoesters. We sought to compare the epimerization levels between the benzaldehyde and nitrophenyl derivatives; however we were unable to synthesize the dipeptide nitophenyl selenoester presumably because of its disposition to undergo azalactonization and form other side products (Scheme 1). The azalactone side product was observed by LCMS of crude reaction mixtures. We hypothesize that the absence of epimerization with the benzaldehyde derivative and its slower rate of aminolysis as compared to the nitrophenyl analog suggests an electronic (n-π^*) interaction between the carbonyl of the formyl group and the selenoester that diminishes azalactone formation.[50-52] Further studies will be required to support this postulate.

Scheme 1



As part of these studies, we compared the reactivity of amino acid selenobenzaldehyde esters with standard activated esters utilized for peptide synthesis. Performance of activated Fmoc-valine esters derived from DCC/DMAP, HBTU, and DCC/NHS (*N*-hydroxysuccinimide) was compared with FmocVal-selenobenzaldehyde ester for condensation with tryptophan methyl ester. Coupling using the classical DCC/DMAP conditions requires 96 hours for 70% conversion. All reactions were performed at 10 mM substrate concentrations in DMF (Supporting Information, Table S1). Reaction with the preformed NHS-ester is completed in 30 h while condensation with HBTU derived OBt ester is completed in 30 mins. Reaction rates for the selenobenzaldehyde ester are slightly slower than for the HBTU-derived OBt ester, a highly efficient but non-chemoselective and water sensitive coupling agent. This analysis suggests that ACL may potentially be useful for performing ligation reactions under aqueous conditions on solid-phase matrices.

Table 3. Scope of the aldehyde capture ligation for the synthesis of peptides. Residues at the junction are underlined.[a]

Entry	Ligation product	Time[b]
1	Fmoc-AW-COOMe	2 min
2	Fmoc-AV-COOMe	2 min
3	Fmoc-AR-COOMe	2 min
4	Fmoc-GC-COOMe	2 min
5	Fmoc-AS-COOBn	5 min
6[c]	$(Fmoc\text{-}A)_2K$-COOH	5 min
7	Fmoc-FL- COO*t*Bu	3 min
8	Fmoc-VW-COOMe	2 h
9	Fmoc-VV-COOMe	4 h
10	Fmoc-AibW-COOMe	2 h
11	Fmoc-AVDE-$CONH_2$	5 min
12	Fmoc-AASY-$CONH_2$	2 min
13[d]	Fmoc-AASY-$CONH_2$	2 min
14	Fmoc-AAAH-$CONH_2$	2 min
15	Fmoc-AVDAFE-$CONH_2$	5 min
16	Fmoc-FVDAFE-$CONH_2$	5 min
17	Fmoc-VAVDAFE-$CONH_2$	5 min

[a]Reaction conditions: Fmoc amino acid-selenobenzaldehyde ester (10 μmol), *N*-terminal amino acid/peptide HCl or CF_3COOH salts (20 μmol) and Et_3N (20 μmol) in 1 mL DMF. [b] Time for > 95 % conversion to product at room temperature. Analysis by HPLC traces of the crude reaction mixture. [c]Double acylation of lysine on the α- and ε-amine groups occurs. [d]FmocAla-selenobenzaldehyde ester (20 μmol), N-terminal CF_3COOH·ASY-$CONH_2$ (10 μmol) and Et_3N (10 μmol) in 1 mL DMF; after 6 h only the amidation product is obtained.

Potential of ACL for chemoselective ligation

The utility and attractiveness of native chemical ligation results from the fact that it affords chemoselective ligation of peptides, and other biomolecules, without need of protecting groups on reactive side chain functionality. We explored the tolerance of

ACL for unprotected side chain groups for a variety of amino acid partners including Arg, Asp, Cys, Glu, His, Leu, Phe, Ser, Trp, Tyr and Val (Table 3). These amino acid residues underwent the desired ligation suggesting that ACL is a chemoselective reaction. Reactions of selenobenzaldehyde esters with N-terminal serine and cysteine residues may provide the amide products through the formation of respective pseudoproline intermediates as reported for oxobenzaldehyde esters,[30,31,39,40] rather than through an ACL mechanism. We investigated this possibility but did not observe the formation of the pseudoproline derivatives with seleno-benzaldehyde esters in DMF or pyridine/acetic acid mixtures (Figure S6). (The serine/threonine ligation has been shown to be optimal in pyridine/acetic acid mixtures.[31]) *N*-terminal cysteine may undergo condensation under an NCL or ACL mechanism.

A potential limitation of ACL is that the aldehyde group on selenobenzaldehyde esters may react with any amine, i.e. ε-amino of lysine would undergo an ACL as well as N-terminal amines (Table 3, entry 6). Possible solutions to control reactions of lysine side chain amines include standard protecting group strategies or reactions with peptides and proteins under pH-controlled conditions. The *N*-terminal amino group of proteins has a significantly lower pK_a value than the ε-amine of lysines.[53] Prior studies have shown that this pK_a difference can be exploited to selectively modify the *N*-terminal amine in the presence of lysine residues.[54] We have explored the possibility of selectively modifying N-terminal amines of proteins using ACL (*vide infra*).

We investigated the possibility if side chain alcohol functionality of serine or tyrosine is acylated, especially with an excess of selenobenzaldehyde esters. Tripeptide ASY-$CONH_2$ was treated with 2 or 0.5 eq of Fmoc-alanine selenobenzaldehyde ester (Table 3, entries 12-13). Acylation of the side chain alcohol group was not observed even after prolonged reaction periods. C-terminal aspartic and glutamic acid thioesters can undergo reactions with side chain carboxylates;[55] this side reaction will likely also preclude placement of Asp and Glu residues, along with other reactive side chains such as cysteine, at the C-terminus in peptido selenobenzaldehydes, as in NCL.

Evaluation of ACL for difficult couplings

The efficiency of peptide fragment couplings by native chemical ligation, and by other coupling agents, is significantly reduced for bulky amino acid substrates.[56] We examined a range of selenobenzaldehyde esters featuring bulky C-terminal residues including valine and aminoisobutyric acid (Aib) to evaluate the potential of ACL for these challenging couplings. The sterically demanding Val and Aib selenobenzaldehyde esters require longer ligation times (2-4 h) for >95 % conversion (Table 3, entries 8-10) as compared to 2-5 minutes for Ala, Gly and Phe analogs (Table 3, entries 1-7).[56] Ligation of selenobenzaldehyde esters with pentapeptide (VDAFE) underwent completion in just 5 minutes (Table 3, entries 15-17), suggesting that ACL with peptides is as efficient as with single amino acid residues.

Solid phase synthesis of peptide selenobenzaldehyde esters: Evaluation of ACL for peptide ligation

To extend the ACL technology to large peptides, we explored a solid phase method for the synthesis of seleno-benzaldehyde esters using the approach outlined by Dawson et al. for the solid phase synthesis of thioesters (Figure 6).[57] A model pentapeptide (FmocLYRAG) N-acyl-benzimidazolinone (Nbz) was synthesized using the Dawson Nbz protocol. The peptide was cleaved as the o-benzaldehydeselenoester by treatment of the resin with diselenide 7 (Figure 6) and tributylphosphine. The crude yield of the cleaved peptide was calculated to be 59% based on resin loading using the Fmoc absorbance. The selenoester was characterized using LCMS (Figure 7a). (The HPLC analysis shows a peak for the hydrolyzed peptide along with the desired selenoester. We posit that the selenoester likely hydrolyzes during the HPLC run under the acidic aqueous conditions utilized for the analysis and not during the selenoester synthesis. The hydrolysed peptide is not observed upon further reaction of the unpurified selenoester with an amine (Figure 7b), supporting this hypothesis).

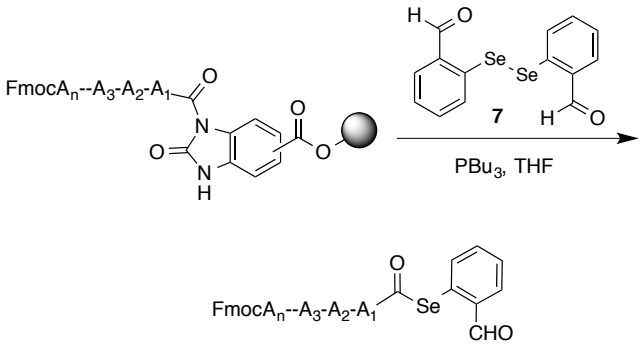


Figure 6. Solid phase synthesis of peptide seleno-*o*-benzaldehyde esters.

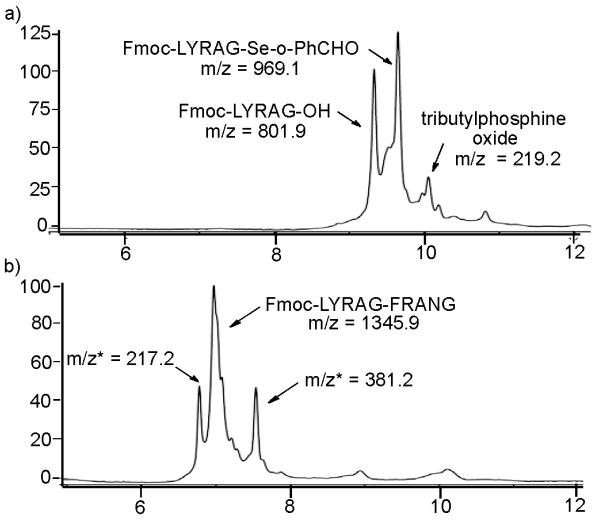


Figure 7. HPLC analysis of peptide-selenoester condensation. (a) Crude analytical HPLC trace of side chain-deprotected FmocLYRAG-Se-*o*-PhCHO synthesized from the corresponding resin bound Nbz-peptide. The peptide was treated with trifluoroacetic acid:TES:H2O (95:2.5:2.5) to remove the protecting groups prior to LCMS analysis. (b) Crude HPLC trace of FmocLYRAGFRANG-$CONH_2$ obtained from the condensation of unpurified FmocLYRAG-Se-*o*-PhCHO from part (a) and FRANG-$CONH_2$. HPLC Conditions: 0.1 % TFA (v/v) in water (solvent A): acetonitrile (solvent B); gradient 5-95% B in 30 min, flow rate = 0.5 mL/min. m/z* = Masses of these peaks do not correspond to the starting selenoester or identifiable side products.

We utilized the solid phase synthesis methodology to evaluate the applicability of ACL to the synthesis of model peptides in aqueous solutions. For direct comparison with NCL, we prepared peptide sequences that have previously been explored by Dawson et al.[56] These coworkers noted that NCL rates of pentapeptides AcLYRAX-SPh, where X is any residue, with CRANK-$CONH_2$ are largely dependent on the identity of X, with bulky residues (i.e. valine) at this position leading to inefficient couplings. We used analogs of these pentapeptide sequences, FmocLYRAG and FmocLYRAV selenobenzaldehyde esters, to probe the effectiveness of ACL for peptide ligations with peptide FRANG-$CONH_2$ (Table 4). As expected, ACL with glycinyl selenobenzaldehyde proceeds much faster than the analogous valinyl selenobenzaldehyde, which requires 8 h. Yet, the ACL reaction rate with FmocLYRAV-Se-*o*-PhCHO is remarkable considering native chemical ligation between AcLYRAV-SPh and CRANK-$CONH_2$ needs more than 48 h for 60% completion.[56] The ACL-mediated condensation of peptides in aqueous solutions is slower than in DMF, requiring roughly three times longer for completion. The HPLC analysis of crude reaction mixture shows efficient conversion of the selenoester to the desired peptide (Figure 7).

Table 4. Aldehyde capture ligation of peptide segments

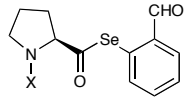
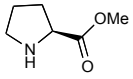
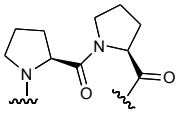
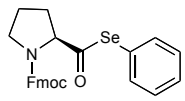
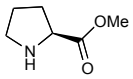
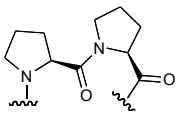
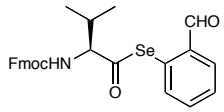
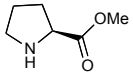
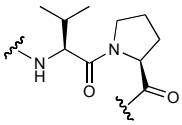
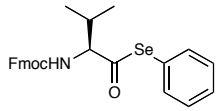
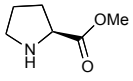
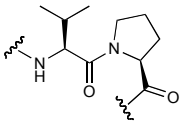
Ligation product[a]	Solvent	Time[b]
Fmoc-LYRAGFRANG-$CONH_2$[c]	DMF	30 min
Fmoc-LYRAGFRANG-$CONH_2$[d]	10% DMF in H_2O	90 min[d]
Fmoc-LYRAVFRANG-$CONH_2$[c]	DMF	8 h

[a]Reaction progress analyzed by HPLC; residues at the junction are underlined. [b]Time required for 95 % conversion to product at room temperature. [c]FmocLYRAX-Se-*o*-PhCHO (2 μmol), FRANG-$CONH_2$ (4 μmol) and Et_3N (4 μmol) in 1 mL DMF. [d]FmocLYRAG-Se-*o*-PhCHO (1 μmol), FRANG-$CONH_2$ (2 μmol) and Et_3N (2 μmol) in 0.5 mL of 100 mM NaH_2PO_4 buffer (pH-6.5):DMF (90:10).

Comparison of ACL and NCL for proline ligation

Acylated prolines thioesters react slowly (<15% yield in >48 h) under NCL conditions.[56] The rates of these NCL reactions have been reported to be independent of the number of residues attached to the N-terminus of proline.[58,59] This observation is consistent with the postulated contribution of the n-π* interaction on proline reactivity as only the carbonyl attached directly to the proline residue perturbs the rate of the reaction.[51] We synthesized various acylated proline derivatives to gauge the potential of proline ligations using the ACL approach (Table 5, entries 1 and 2), and find that ACL excels at difficult ligation of proline residues indiscriminant of whether it is a C-terminal or N-terminal residue. Couplings of FmocVal, FmocPro, FmocAla-Pro and FmocVal-Pro selenobenzaldehyde esters with proline methyl ester proceed rapidly (Table 5, entries 1 and 3); whereas, analogous reactions of FmocVal and FmocPro selenophenyl esters (without the *ortho*-aldehyde group) require significantly longer periods (Table 5, entries 2 and 4). These studies demonstrate that preformed phenylselenoesters, which are typically considered to be highly reactive,[35] are less efficient than the ACL auxiliary for difficult couplings.

Table 5. Coupling of Proline Residues with ACL.[a]

Entry	C-terminus	N-terminus	Ligation site	Conversion (%)[b]	Time[c]
1	 X = Fmoc, P1[d], P2			> 95	30 min
2[e]				~ 20	7 h
3				> 95	2 h
4				> 95	> 96 h

[a]Reaction conditions: C-terminal seleno-benzaldehyde ester (10 μmol), N-terminal AA/peptide (20 μmol) and Et_3N (20 μmol) in 1 mL DMF. [b]Analysis by HPLC traces of the crude reaction mixture. [c]Room temperature. [d]Reaction in buffer (pH-8.5) > 95 % conversion in 90 min. P1: Fmoc-A––, P2: Fmoc-V––. [e]After 10 h, unidentified side products are observed.

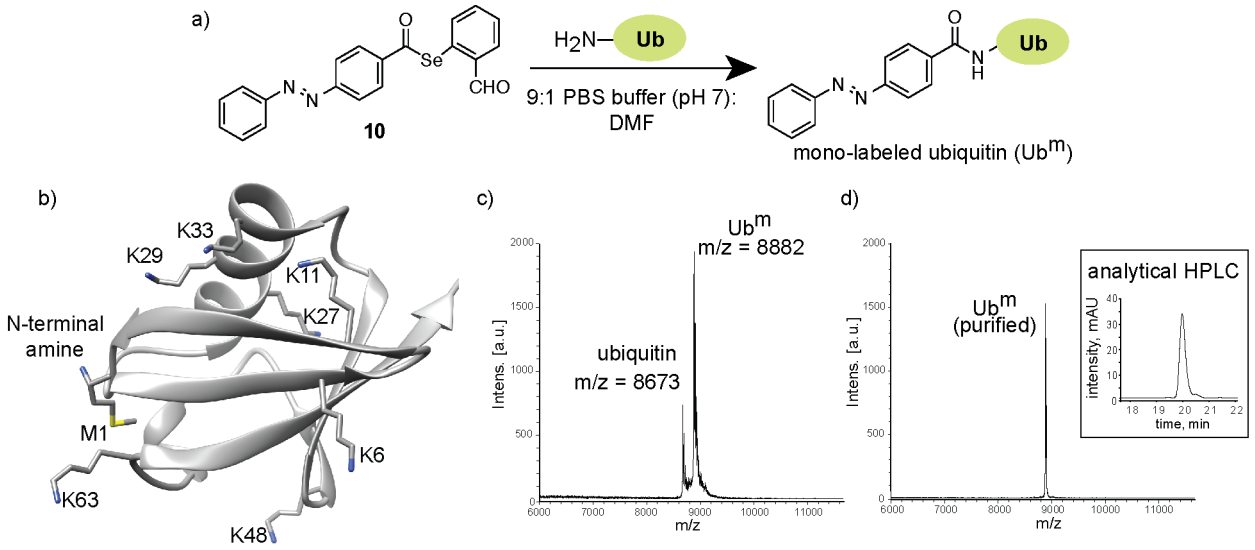


Figure 8. (a) Ubiquitin (^{15}N labelled) was treated with selenoester **10** in 10% DMF in 1X phosphate buffered saline (PBS), pH 7.0. Mono-labeling of ubiquitin was observed as the exclusive product. (b) Ubiquitin contains seven lysine residues and provides a stringent test for evaluation of the specificity of reaction at the *N*-terminus. (c) The reaction progress was evaluated by MALDI and LCMS; the MALDI spectrum after 40 hours is shown. MS/MS analysis confirms that labelling is localized to the *N*-terminus (Figure S7). (d) MALDI MS of the HPLC purified sample. The analytical HPLC trace is shown in the inset.

Potential of ACL for protein modifications: N-terminal modification of ubiquitin as a model protein

ACL adds to a growing list of methods that allow chemoselective formation of the amide bond in aqueous buffers. The attractiveness of ACL is that it potentially allows coupling of any set of amino acid residues, without requiring specific *N*-terminal amino acids such as cysteine. A potentially significant limitation is that the aldehyde may capture any amine, i.e., it may be difficult to selectively modify lysine side chains or the *N*-terminal amines. Two solutions may be envisioned to address this potential limitation: (a) in short peptides or other synthetic oligomers and small molecules, the appropriate amine may be protected using standard approaches; (b) the pKa difference between the *N*-terminal amino group and the ε-amine of lysine may be exploited to achieve selective modification of the *N*-terminus.[53,54,60]

We evaluated the potential of ACL to specifically acylate the *N*-terminus of ubiquitin. We conjectured that ubiquitin would provide a stringent test to determine the specificity of reaction at the *N*-terminus because it contains seven lysine residues (Figure 8). We used diazobenzene dye **10** as a model substrate for reaction with ubiquitin. Five-fold excess of the dye selenoester was incubated with the protein (100 μM) in 10% DMF in 1X PBS buffer, pH 7.0. The reaction was monitored using mass spectroscopy (MALDI and LCMS). After 40 h, roughly 70% of ubiquitin was modified by the dye and only the peak corresponding to mono-labeled ubiquitin (Ub^m) was observed by Maldi and LCMS (Figure 8). Further analysis by MS/MS showed that the modification is localized to the *N*-terminal hexapeptide fragment of ubiquitin (Figure S7). This result further highlights the attractiveness of the ligation strategy proposed herein, and suggests potential uses of ACL as a strategy for *N*-terminal modification of proteins.[61-63]

In summary, we have introduced a new method for peptide ligation, which builds on the fundamental principles utilized by common ligation techniques. The key feature of aldehyde capture ligation is that it employs the rapid association between an aldehyde group and an amine to enforce an intramolecular reaction leading to the desired native amide bond formation. Because primary and secondary amines react with the selenobenzaldehyde esters readily, we postulate that the hemiaminal II (Figure 1) is the active intermediate in the reaction, but contributions from the imine as a reactive intermediate with primary amines or direct acylation of selenoesters cannot be ruled out. We have demonstrated the potential of this approach for ligating a variety of unprotected amino acids and peptides, including difficult sequences. Preliminary results indicating mono-labeling of ubiquitin, a model protein, highlight the potential of ACL as a strategy for protein modification. Importantly, we outline a strategy for the facile synthesis of peptide selenoesters on solid phase, which will further enable evaluation of the ACL methodology for the synthesis of more complex peptides.

EXPERIMENTAL SECTION

General. Commercial-grade reagents and solvents were used without further purification except as indicated. All reactions were stirred magnetically or mechanically shaken; moisture-sensitive reactions were performed under nitrogen or argon atmosphere. Reverse-phase HPLC experiments were conducted with 0.1% aqueous trifluoroacetic acid and acetonitrile as eluents with 4.6 mm × 150 mm (analytical scale) or 21.4 mm × 150 mm (preparative scale) Waters C18 Sunfire columns using a Beckman Coulter HPLC equipped with a System Gold 168 Diode array detector. The typical flow rates for analytical and preparative HPLC were 0.5-1 mL/min and 8 mL/min, respectively. ESI-MS data were obtained on an Agilent 1100 series LC/MSD (XCT) electrospray trap. Protein MALDI data was collected on Bruker MALDI-TOF/TOF UltrafleXtreme Spectrometer. MS/MS

analysis was performed on Agilent 1100 Series LCMSD VL MS Spectrometer. Proton NMR spectra of selenoesters were recorded on a Bruker AVANCE 400 MHz spectrometer.

Peptide synthesis. Peptides were synthesized on a CEM Liberty microwave peptide synthesizer using Fmoc solid-phase chemistry on Rink amide resin and purified by reversed-phase HPLC. The identity and the purity of the peptides were confirmed by ESI-MS.

General procedure for synthesis of thiobenzaldehyde and selenobenzaldehyde esters: Amino acid (AA)/peptide (1 mmol) was dissolved in anhydrous THF (5 mL) in an oven-dried round bottom flask under N_2. The resulting solution was stirred and cooled at 0 °C in ice-bath for 10 min. Then, DCC (1 mmol) was added at 0 °C. The reaction was stirred at 0 °C for 30 min. Then, di-thio (**6**) or di-seleno benzaldehyde (**7**) (0.5 mmol), TCEP·HCl (0.6 mmol), Et_3N (0.6 mmol) and 2 drops of water were added. The resulting reaction mixture was stirred at room temperature for 30 min. The reaction was then cooled in ice bath and filtered to remove dicyclohexyl urea. The filtrate was concentrated under vacuum and purified by flash chromatography or HPLC to give AA/peptide-thio-benzaldehyde or AA/peptide-seleno-benzaldehyde ester, respectively as pale yellow solid (55-70% yield).

General procedure for the Aldehyde Capture Ligation: To a solution of C-terminal amino acid or peptide-oxo, thio or seleno ester (2-10 μmol) in 1 mL DMF, was added N-terminal amino acid or peptide (2 eq) and Et_3N (2 eq). The reaction was stirred at 22 °C.

Analytical analysis of ACL reactions: For determination of reaction rates, the reaction mixture was diluted 20-fold in acetonitrile and kept over dry-ice until HPLC analysis. HPLC (C-18 columns): 0.1% TFA (v/v) in water (solvent A): acetonitrile (solvent B); gradient 45-85 % in 30 min, flow rate = 0.5 mL/min, detection wavelength 280 nm.

ASSOCIATED CONTENT

Supporting Information.
Supporting Figures and experimental procedures and analytical data for new compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

arora@nyu.edu

Present Addresses

†Department of Chemistry & Biochemistry, Seton Hall University, South Orange, NJ 07079.

Author Contributions

‡These authors contributed equally to this work.

ACKNOWLEDGMENTS

We thank the US National Science Foundation (CHE-1506854) for funding. P.S.A. thanks Anna Mapp, Laura Kiessling and Neville Kallenbach for helpful suggestions.

REFERENCES

(1) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249.
(2) Davis, B. G. *Science* **2004**, *303*, 480.
(3) Nilsson, B. L.; Soellner, M. B.; Raines, R. T. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91.
(4) Pattabiraman, V. R.; Bode, J. W. *Nature* **2011**, *480*, 471.
(5) Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chem. Int. Ed.* **2008**, *47*, 10030.
(6) Stephanopoulos, N.; Francis, M. B. *Nat. Chem. Biol.* **2011**, *7*, 876.
(7) Ogunkoya, A. O.; Pattabiraman, V. R.; Bode, J. W. *Angew. Chem. Int. Ed.* **2012**, *51*, 9693.
(8) Tam, J. P.; Xu, J.; Eom, K. D. *Peptide Science* **2001**, *60*, 194.
(9) Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 2670.
(10) Wang, P.; Dong, S.; Shieh, J.-H.; Peguero, E.; Hendrickson, R.; Moore, M. A. S.; Danishefsky, S. J. *Science* **2013**, *342*, 1357.
(11) Payne, R. J.; Wong, C.-H. *Chem. Commun.* **2010**, *46*, 21.
(12) Scheck, R. A.; Francis, M. B. *ACS Chem. Biol.* **2007**, *2*, 247.
(13) Gamblin, D. P.; Scanlan, E. M.; Davis, B. G. *Chem. Rev.* **2008**, *109*, 131.
(14) Chalker, J. M.; Bernardes, G. J. L.; Davis, B. G. *Acc. Chem. Res.* **2011**, *44*, 730.
(15) Kiessling, L. L.; Splain, R. A. *Annu. Rev. Biochem* **2010**, *79*, 619.
(16) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem* **2000**, *69*, 923.
(17) Raibaut, L.; Ollivier, N.; Melnyk, O. *Chem. Soc. Rev.* **2012**, *41*, 7001.
(18) Bode, J. W.; Fox, R. M.; Baucom, K. D. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 1248.
(19) Wang, T.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134*, 13244.
(20) Shen, B.; Makley, D. M.; Johnston, J. N. *Nature* **2010**, *465*, 1027.
(21) Noda, H.; Eros, G.; Bode, J. W. *J. Am. Chem. Soc.* **2014**, *136*, 5611.
(22) Pattabiraman, V. R.; Ogunkoya, A. O.; Bode, J. W. *Angew. Chem. Int. Ed.* **2012**, *51*, 5114.
(23) Aimoto, S. *Peptide Sci.* **1999**, *51*, 247.
(24) Payne, R. J.; Ficht, S.; Greenberg, W. A.; Wong, C.-H. *Angew. Chem. Int. Ed.* **2008**, *47*, 4411.
(25) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776.
(26) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. *Org. Lett.* **2000**, *2*, 2141.
(27) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2000**, *2*, 1939.
(28) Kemp, D. S. *Biopolymers* **1981**, *20*, 1793.
(29) Coltart, D. M. *Tetrahedron* **2000**, *56*, 3449.
(30) Li, X.; Lam, H. Y.; Zhang, Y.; Chan, C. K. *Org. Lett.* **2010**, *12*, 1724.
(31) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6657.
(32) Kemp, D. S.; Vellaccio, F., Jr. *J. Org. Chem.* **1975**, *40*, 3003.
(33) Kemp, D. S.; Grattan, J. A.; Reczek, J. *J. Org. Chem.* **1975**, *40*, 3465.
(34) Ishiwata, A.; Ichiyanagi, T.; Takatani, M.; Ito, Y. *Tetrahedron Lett.* **2003**, *44*, 3187.
(35) Durek, T.; Alewood, P. F. *Angew. Chem. Int. Ed.* **2011**, *50*, 12042.
(36) McGrath, N. A.; Raines, R. T. *Acc. Chem. Res.* **2011**, *44*, 752.
(37) Mautner, H. G.; Chu, S.-H.; Gunther, W. H. H. *J. Am. Chem. Soc.* **1963**, *85*, 3458.
(38) Chu, S.-H.; Mautner, H. G. *J. Org. Chem.* **1966**, *31*, 308.
(39) Liu, C.-F.; Tam, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 4149.
(40) Liu, C. F.; Tam, J. P. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6584.
(41) Holmquist, B.; Bruice, T. C. *J. Am. Chem. Soc.* **1969**, *91*, 2985.

(42) Jencks, W. P.; Gilchrist, M. *J. Am. Chem. Soc.* **1968**, *90*, 2622.
(43) Trmcic, M.; Hodgson, D. R. *Beilstein J. Org. Chem.* **2010**, *6*, 732.
(44) Yang, W.; Drueckhammer, D. G. *Org. Lett.* **2000**, *2*, 4133.
(45) Castro, E. A. *Chem. Rev.* **1999**, *99*, 3505.
(46) Bodanszky, M. *Nature* **1955**, *175*, 685.
(47) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; McGraw-Hill: New York, 1969.
(48) Syper, L.; Mlochowski, J. *Tetrahedron* **1988**, *44*, 6119.
(49) Bodanszky, M. In *Principles of Peptide Synthesis*; Springer Berlin Heidelberg: 1993, p 169.
(50) Hinderaker, M. P.; Raines, R. T. *Protein Sci.* **2003**, *12*, 1188.
(51) Choudhary, A.; Raines, R. T. *Protein Sci.* **2011**, *20*, 1077.
(52) Hodges, J. A.; Raines, R. T. *Org. Lett.* **2006**, *8*, 4695.
(53) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Sci.* **2009**, *18*, 247.
(54) Chan, A. O.-Y.; Ho, C.-M.; Chong, H.-C.; Leung, Y.-C.; Huang, J.-S.; Wong, M.-K.; Che, C.-M. *J. Am. Chem. Soc.* **2012**, *134*, 2589.
(55) Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, *2003*, 3267.
(56) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068.
(57) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2008**, *47*, 6851.
(58) Pollock, S. B.; Kent, S. B. H. *Chem. Commun.* **2011**, *47*, 2342.
(59) Townsend, S. D.; Tan, Z.; Dong, S.; Shang, S.; Brailsford, J. A.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134*, 3912.
(60) Bernal-Perez, L. F.; Prokai, L.; Ryu, Y. *Anal. Biochem.* **2012**, *428*, 13.
(61) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 5307.
(62) Theile, C. S.; Witte, M. D.; Blom, A. E. M.; Kundrat, L.; Ploegh, H. L.; Guimaraes, C. P. *Nat. Protocols* **2013**, *8*, 1800.
(63) Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. *Angew. Chem. Int. Ed.* **2012**, *51*, 9377.

Table of Content: