Published in final edited form as:

J Am Chem Soc. 2016 September 14; 138(36): 11775–11782. doi:10.1021/jacs.6b05719.

A Helping Hand to Overcome Solubility Challenges in Chemical Protein Synthesis

Michael T. Jacobsen^{†,||}, Mark E. Petersen[†], Xiang Ye[‡], Mathieu Galibert[§], George H. Lorimer[‡], Vincent Aucagne^{§,*}, and Michael S. Kay^{†,*}

[†]Department of Biochemistry, University of Utah School of Medicine, 15 N Medical Drive East Rm 4100, Salt Lake City, Utah 84112-5650, United States

[‡]Department of Chemistry & Biochemistry, 8051 Regents Drive, University of Maryland, College Park, Maryland, 20742-4454, United States

[§]Centre de Biophysique Moléculaire, CNRS UPR4301, Rue Charles Sadron, 45071 Orléans Cedex 2, France

Abstract

Although native chemical ligation (NCL) and related chemoselective ligation approaches provide an elegant method to stitch together unprotected peptides, the handling and purification of insoluble and aggregation-prone peptides and assembly intermediates create a bottleneck to routinely preparing large proteins by completely synthetic means. In this work, we introduce a general new tool, Fmoc-Ddae-OH, *N*-Fmoc-1-(4,4-dimethyl-2,6-dioxocyclo-hexylidene)-3-[2-(2-aminoethoxy)ethoxy]-propan-1-ol, a hetero-bifunctional traceless linker for temporarily attaching highly solubilizing peptide sequences (*helping hands*) onto insoluble peptides. This tool is implemented in three simple and nearly quantitative steps: (i) on-resin incorporation of the linker at a Lys residue ϵ -amine, (ii) Fmoc-SPPS elongation of a desired solubilizing sequence, and (iii) in-solution removal of the solubilizing sequence using mild aqueous hydrazine to cleave the Ddae linker after NCL-based assembly. Successful introduction and removal of a Lys₆ helping hand is first demonstrated in two model systems (Ebola virus C20 peptide and the 70-residue ribosomal protein L31). It is then applied to the challenging chemical synthesis of the 97-residue co-chaperonin GroES, which contains a highly insoluble C-terminal segment that is rescued by a helping hand. Importantly, the Ddae linker can be cleaved in one pot following NCL or desulfurization. The purity, structure, and chaperone activity of synthetic L-GroES were validated with respect to a recombinant control. Additionally, the helping hand enabled synthesis of D-GroES, which was inactive in a hetero-chiral mixture with recombinant GroEL—providing additional insight into chaperone specificity. Ultimately, this simple, robust, and easy-to-use tool is expected to be broadly applicable for the synthesis of challenging peptides and proteins.

*Corresponding Authors: aucagne@cns-orleans.fr (V.A.), kay@biochem.utah.edu (M.S.K.).

^{||}Present Addresses

Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, U.K.

M.S.K. is a Scientific Director, consultant, and equity holder of the D-Peptide Research Division of Navigen, which is commercializing D-peptide inhibitors of viral entry.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Graphical Abstract

Keywords

native chemical ligation; ligation; protein chemical synthesis; peptide synthesis; linker; solid-phase synthesis; protein folding; chaperones

INTRODUCTION

The development of native chemical ligation (NCL)^{1,2} by Kent's group was a seminal discovery for the total chemical synthesis of peptides and proteins. Using NCL, peptides containing N-terminal Cys and C-terminal thioesters can be chemoselectively ligated to generate larger synthetic products.^{2,3} This synthetic strategy—solid-phase peptide synthesis (SPPS) to prepare peptide segments followed by NCL or another ligation method^{4–6} to stitch the peptides together—has been very successful for producing synthetic proteins to probe biological problems that would be difficult to assess using molecular biology. Notable achievements include innovative studies on ubiquitination pathways,^{7–10} extraordinary efforts to prepare well-defined glycoforms of erythropoietin,^{11–18} and total syntheses of mirror-image proteins not accessible by recombinant means^{19–21}—among many other interesting targets (recently reviewed^{22,23}). However, total chemical syntheses of these protein targets can be unexpectedly challenging, due to difficulties in handling (dissolving, purifying, and reacting) poorly soluble and aggregation-prone sequences. In particular, certain hydrophobic peptide segments, prominent in membrane proteins, remain especially challenging.^{24–26} Furthermore, it would be highly beneficial to deploy a tool for routinely increasing peptide/protein concentrations to the mM range in order to generally accelerate chemoselective ligation reactions.^{27,28}

To address this challenge, many groups have devised chemistries for temporarily improving peptide segment solubility. One of the most prominent examples is the thioester poly-Arg tag, introduced by the Kent²⁹ and Aimoto³⁰ groups. In this method, the peptide segment of interest is directly synthesized and purified with a highly solubilizing peptide sequence (Arg_n, where n = 5–9) introduced on the thioester-leaving group.

This method has been used in several cases^{31–33} to endow peptides with greatly improved solubility and handling properties. However, the solubilizing tag is eliminated during the ligation reaction (via *in situ* trans-thioesterification with a better thiol leaving group), so the

solubility enhancement cannot be retained during or beyond the ligation step. Furthermore, these tags are difficult to install by Fmoc-SPPS due to the instability of thioesters to piperidine-mediated Fmoc deprotection conditions,³⁴ although an indirect approach using ortho-aminoanilide cryptothioesters has been described.³⁵ A less commonly used route for introducing semi-permanent C-terminal solubilizing tags entails C-terminal base-labile linkers—(e.g. esters) introduced either by Boc³⁶ (glycolic acid linker) or Fmoc-SPPS^{37–39} (4-hydroxymethylbenzoic acid linker). An approach using the acid-labile phenylacetamido (PAM) linker has also been described⁴⁰. Alternatively, linkers containing substrate sequences cleaved by HIV-protease³² and carboxypeptidase⁴¹ have also been used. Thus far, these base-, acid- and enzyme-sensitive linkers have been limited to peptide termini.

Some versatile approaches to this problem have recently been developed. Danishefsky's group designed an allyl-based protection method for introducing semi-permanent solubilizing groups by preparing custom Fmoc-Glu and Fmoc-Lys building blocks linked to solubilizing guanidinium moieties through ester and carbamate groups, respectively.⁴² Liu's group developed a photolabile linker for incorporating semi-permanent solubilizing sequences at Fmoc-Gln building blocks.⁴³ The same group also introduced an elegant two-step method for selectively removing semi-permanent solubilizing groups at Fmoc-Gly building blocks.⁴⁴ Here, a TFA-stable derivative of the *N*-(2-hydroxy-4-methoxy benzyl) (Hmb) group⁴⁵ is converted *in situ* into a TFA-labile Hmb during the NCL reaction.^{46,47} Building on this work, Liu's group recently published a method to introduce removable backbone modifications (RBMs) via a salicylaldehyde derivative.⁴⁸

A New Helping Hand Approach

Unfortunately, no single method for introducing tailor-made semi-permanent solubilizing sequences has been widely accepted in the field for multiple reasons, including complex reagent synthesis, potentially damaging cleavage conditions, slow kinetics of incorporation or elimination, limited placement within peptide sequences, or lability to NCL conditions. We desire a general chemical tool for routinely introducing semi-permanent solubilizing sequences (“*helping hands*”) onto difficult peptides. Accordingly, we had five requirements:

1. Compatibility and facile incorporation by Fmoc-SPPS,
2. Compatibility with NCL and other chemical conditions typically employed in assembly of peptide segments (including desulfurization and oxidative hydrazide activation),
3. Site-selective attachment of the designed solubilizing sequence,
4. Ability to be incorporated at several sites in nearly all peptide segments of the protein target (i.e., not limited to N- or C-termini),
5. Simple and selective removal of the linker to generate native peptide after the solubilizing application is complete.

In response to this challenge, we describe here synthesis and applications of Fmoc-Ddae-OH, *N*-Fmoc-1-(4,4-dimethyl-2,6-dioxocyclo-hexylidene)-3-[2-(2-aminoethoxy)ethoxy]-propan-1-ol, a new hetero-bifunctional traceless linker for attaching solubilizing groups to difficult peptide sequences (Figure 1a). With this simple-to-prepare tool, we can attach and

later remove highly solubilizing peptide sequences on any peptide segment containing a Lys residue, via on-resin Fmoc-SPPS elongation and aqueous hydrazine treatment, respectively. To validate the versatility and robustness of our tool, we applied it in three different synthetic contexts: a model 20-residue peptide and two proteins of 70 and 97 residues requiring NCL-based assembly. In particular, we prepared both L- and D-versions of the synthetically challenging heptameric chaperone protein GroES.

RESULTS AND DISCUSSION

Design and Synthesis of Fmoc-Ddae-OH Linker

The inspiration for our linker comes from Bycroft's pioneering work^{49–51} on the (4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl (Dde) amine protecting group. The enamine bond is selectively cleaved by treatment with bis- α -nucleophiles such as hydrazine or hydroxylamine⁵² (as well as NaBH_4 ⁵³), but is stable to Fmoc elongation and deprotection conditions. Due to its robustness and specific cleavage conditions, a number of Dde-based linkers have been developed for various applications, including SPPS,⁵⁴ oligosaccharide synthesis,⁵⁵ polyamine synthesis^{49,56,57}, affinity purification,⁵⁸ and protein labeling.⁵⁹

We prepared Fmoc-Ddae-OH (Figure 1a) in one step, by reaction of commercially available *N*-Fmoc-amido-PEG₂-acid with dimedone (Figures S1–S4). The general strategy for implementing the new linker is shown in Figure 1b. Here, a “difficult” peptide is prepared by standard Fmoc-SPPS, incorporating a single internal Lys residue with an orthogonal protecting group (Dde in this case), while the N-terminus is protected by a piperidine-stable group (Boc in this case) (Figure 1b, Step 1). Upon elongation of the linear peptide chain, but before cleavage from the resin, the Dde group is selectively removed with 3% hydrazine in DMF to reveal a single primary amine in the solid-supported peptide (Figure 1b, Step 2).

Fmoc-Ddae-OH can then be directly reacted (no other additives are needed, and excess reagent can be recycled) with the primary amine to introduce a fresh *N*-Fmoc protected amine linked to the peptide sequence through a hydrophilic (PEG₂) Dde-derived cleavable moiety (Figure 1b, Step 3). Next, Fmoc-SPPS can be used to build a solubilizing peptide sequence (helping hand) (Figure 1b, Step 4) on the Lys side chain. Note that we previously observed partial instability under aqueous conditions of a Dde-based linker on the N-terminus, while Dde linked on Lys side chain was stable over 24 h.⁶⁰ Cleavage from resin and HPLC purification can then be performed to obtain the purified peptide segment equipped with a solubilizing tag (Figure 1b, Step 5).

After NCL-based protein assembly, cleavage of the linker and restoration of native peptide structure can then be performed by addition of aqueous hydrazine (Figure 1b, Step 6), mild conditions similar to those used for acetate deprotection during glycopeptide and glycoprotein synthesis.^{61,62}

Ddae Linker Introduction, Stability and Cleavage in a Model Peptide

A model peptide, Ebolavirus C20 (Ac-DWTKNITDKIDQIIHDFVDK-NH₂), was used to characterize incorporation and removal of the Ddae linker with a Lys₆ helping hand. The

C20 peptide is derived from the C-terminal heptad repeat region of the Ebolavirus GP2 protein.^{63,64}

We first synthesized C20 (Figure 2a) including a central Lys(Dde) (Figure 2b, S15–S17), then removed the Dde group using hydrazine (Figure 2c, S18–S20). Fmoc-Ddae-OH (200 mM in NMP) was then added to the peptide resin for seven hours to generate C20(Fmoc-Ddae) (Figure 2d, time course studies show nearly complete reaction after two h, Figure S21–S23). On-resin Fmoc-SPPS was next performed to extend a Lys₆ helping hand, followed by cleavage from resin using standard TFA conditions to generate C20(K₆-Ddae) (Figure 2e, S24–S26). The cleanness of the crude products (Figures 2b–e, S27) highlights the efficiency of the multistep process.

We next evaluated removal of the helping hand in solution via hydrazine treatment by conducting a time-course study to measure the conversion of C20(K₆-Ddae) to C20. Purified C20(K₆-Ddae) was dissolved in 6 M GuHCl, 100 mM phosphate buffer, pH 7.5. Cleavage of the K₆-Ddae group was triggered by addition of a hydrazine solution in phosphate buffer (final 1 M hydrazine, pH 7.5) (see section S8 of SI for specific details on preparing the hydrazine solution). Clean elimination of the (K₆-Ddae) group was achieved with a $t_{1/2}$ of ~1 h (Figure 2f, S28–31). This C20 peptide, prepared using a helping hand approach, possessed matched HPLC and MS traces with C20 prepared without Fmoc-Ddae linker (compare Figures 2c, S18–20 with Figures 2f, S31–S32).

We also examined the stability of C20(K₆-Ddae) under several different conditions employed in modern chemical peptide/protein synthesis including acidic HPLC buffer (Figure S33), denaturing conditions at low (Figure S34, S36) and neutral pH (Figure S35), and a typical NCL buffer (Fig S37). No significant Ddae cleavage was observed under these standard conditions (Figure S39). In contrast, we did observe partial sensitivity of the linker (~20% cleavage after 24 h) to thiazolidine ring-opening conditions^{65–67} (Figure S38–S39). However, we tested relatively harsh conditions, 200 mM MeONH₂ for 24 h, while most reports employ either lower MeONH₂ concentrations (20–50 mM) or shorter incubation times (3–5 h).

Our studies with the C20 model peptide show that the Fmoc-Ddae-OH linker can be used to introduce solubilizing sequences at Lys side chains, while quick and clean removal of the linker is achieved with 1 M hydrazine. Importantly, the linker is stable to several buffers commonly used in chemical peptide/protein synthesis.

NCL-based Synthesis of Ribosomal Protein L31 Using a Helping Hand

We next pursued the synthesis of a full-length protein target via NCL. Here, we selected the L31 protein from *E. coli*. L31 (Uniprot P0A7M9, RL31_ECOLI) is a 70-residue protein within the large ribosomal subunit.^{68,69}

The synthesis strategy for preparing L31 entailed three peptide segments (L31-1, 2 and 3), including Lys²³ in segment 2 modified with a Lys₆ helping hand (Figure 3a–c). All three starting peptide segments were synthesized and purified (Figure 3d, S40–S51). L31-1 and L31-2 were prepared as peptide hydrazides for subsequent *in situ* activation/thiolysis for

NCL,^{70–73} while L31-3 was prepared with a C-terminal carboxylic acid, in order to match the natural protein (Figure 3c).

The first ligation reaction, L31-1 to L31-2 at the Ser¹⁵-Cys¹⁶ junction, was complete in 1 h (Figure S52). The ligation product L31-1-2 was then purified by HPLC—confirming that the helping hand was stable in an NCL reaction (Figure 3e, S53–S55). We next ligated L31-1-2 to L31-3 to directly evaluate the linker's stability to oxidative hydrazide activation conditions. However, during preliminary studies of this reaction, we encountered significant lactam formation, which can be a challenge with Lys thioesters.⁷⁴ In this case, we overcame the problem by activating L31-1-2 hydrazide in batches that were sequentially added to L31-3 in the ligation reaction (see SI for details). In addition to demonstrating that the second ligation was successful (Figure S56–S57), we were able to purify both the MESNa thioester of L31-1-2 (Figure S58–S60) and the full-length product with helping hand intact (Figure S61–S63). Critically, the linker was stable to peptide hydrazide activation conditions (20 mM NaNO₂, pH 3, –20 °C, 20 min).^{70–73}

We next showed that the helping hand could be removed in one pot after ligation of the full-length product by treatment with 1 M hydrazine, pH 7.5 for 2.5 h (Figure S64). We generated clean and correct full-length product, without helping hand, using this straightforward one-pot ligation and helping hand removal strategy (Figure 3f, S65–S67).

Synthesis Strategy and Initial Efforts to Synthesize GroES

Having validated the helping hand approach in Ebolavirus C20 peptide and *E. coli* L31 protein, we pursued a more ambitious and biologically relevant protein target: the *E. coli* co-chaperonin GroES. Our interest in synthesizing GroES was motivated by our previous work showing that the Gro-EL/ES chaperone was capable of folding both natural (L–) and mirror-image (D–) forms of a synthetic substrate protein, DapA.⁷⁵ This result demonstrated that GroEL/ES recognizes substrate proteins via nonspecific (i.e., non-stereoselective) hydrophobic interactions.

We were interested to probe this interaction more deeply. On this note, crystal structures of GroEL/ES complexes have shown considerable structural heterogeneity in the interaction of the GroES mobile loop with GroEL.⁷⁶ Another group reported that substitution of six positions in the GroES mobile loop with Ala residues did not perturb GroEL complementation in *E. coli* cells.⁷⁷ Additionally, Gierasch's lab showed by NMR that both L- and D-peptides could bind to GroEL.⁷⁸ These data suggest a high degree of plasticity in the GroEL/ES interface. Based on this information and our previous data, we hypothesized that a hetero-chiral complex (L-GroEL/D-GroES or D- GroEL/L-GroES) may possess chaperone activity. To test this hypothesis, we pursued the chemical synthesis of D- GroES (97 residues) to be tested in complex with recombinant (L–) GroEL (548 residues). We also pursued the chemical synthesis of L-GroES as a positive control to validate the activity of our synthetic material.

Total chemical synthesis of *E. coli* GroES (Uniprot P0A6F9, CH10_ECOLI) was reported in 1991 by an ambitious preparation of the entire protein via Boc-SPPS; however, no analytical HPLC or MS data were provided on the synthetic material.⁷⁹ We reasoned that modern

advances in peptide synthesis and ligation technologies made this protein well suited for a multi-segment synthetic approach.

Our retrosynthetic strategy involved two peptide segments, with a single ligation junction at position Leu⁴¹-Ala⁴² (Figure 4a, b). Following ligation of the GroES-N and GroES-C peptides, desulfurization (Cys→Ala) at position 42 would generate full-length native GroES. Unfortunately, our efforts to prepare GroES in this manner failed due to severe handling difficulties with the GroES-C peptide. Specifically, we were unable to dissolve a significant amount of peptide in any usable solvent (e.g., water/acetonitrile + 0.1% TFA, 6 M GuHCl + 5% AcOH, DMF, and ammonium acetate/water/acetonitrile pH 6), nor could we isolate a clean product by RP-HPLC, Fig S68.

These challenges with GroES-C forced us to re-think our synthesis strategy. We considered introducing an additional ligation site C-terminal to Leu⁴¹ in order to break up this difficult segment. However, the absence of any Cys or Ala residues in this region (residues 43 – 97) complicated this prospect. Potentially, a non-natural thiolated residue (reviewed^{28,80,81}) could be used in this region. However, this would require multi-step chemical syntheses of both L- and D-versions of a specialty thiolated amino acid—as well as extra handling steps (additional ligation and desulfurization) in the GroES assembly. Furthermore, we found that a shorter version of GroES-C (peptide 61–97) was even less soluble and could not be dissolved for HPLC analysis.

Synthesis of L- and D-GroES Using a Helping Hand

Based on these challenges, we incorporated a Lys₆ helping hand in the GroES-C peptide (at Lys⁷⁷, Fig 4b,c). The four starting peptide segments for generating L- and D-GroES were then synthesized: L- and D-GroES-N being prepared as peptide hydrazides, while L-GroES-C and D-GroES-C were prepared as C-terminal acids to match the natural protein. Standard Fmoc building blocks and SPPS conditions were used in the syntheses, with a few exceptions. First, several Xaa-Ser(Ψ^{Me,Me}Pro) pseudoproline dipeptides were utilized in all four peptides to enhance synthesis quality by reducing peptide aggregation during SPPS (underlined, Figure 4b). The non-commercially available D-pseudoproline dipeptides were synthesized according to Mutter's original protocol, with slight modifications (Figures S5–S14).^{82,83} Additionally, achiral (Dmb)Gly was used at Gly⁷⁰ to prevent piperidine-mediated aspartimide formation at the Asp⁶⁹-Gly⁷⁰ site. Lys(Dde) was used at position Lys⁷⁷ to install the helping hands, while Boc-Cys(Trt) was incorporated at position 42 to protect the GroES-C N-terminus. Upon completion of the linear GroES-C peptide sequences, the Dde group was removed on-resin, Fmoc-Ddae-OH was then coupled, followed by Fmoc-SPPS to generate a Lys₆ helping hand. All four peptides (L-, D-GroES-N, and L-, D-GroES-C) were then cleaved, purified by preparative HPLC, and characterized by HPLC and LC/MS (Figures S69–S76, S84–S91).

Peptide hydrazides L- and D-GroES-N were activated via NaNO₂ oxidation and MPAA thiolysis to generate thioesters for reaction with their partner Cys-peptides L- and D-GroES-C, respectively. Gratifyingly, ligations were complete in less than 4 h, showing no noticeable traces of products arising from premature Ddae cleavage or related undesirable reactions. The ligation products were purified by HPLC and characterized by analytical HPLC and

LC/MS (Figures S77–S80, S92–S95). We obtained isolated yields of 24 and 21% for L- and D-GroES-NC based on limiting GroES-C peptides.

We next proceeded to the final synthetic step: one-pot desulfurization and helping hand removal to generate full-length, native L- and D-GroES. Peptides L- and D- GroES-NC(Cys⁴², K₆-Ddae⁷⁷) were desulfurized⁸⁴ using free radical method⁸⁵ to convert the ligation junction Cys⁴² into Ala⁴² (Figures S81, S96). The desulfurization was complete in less than 4 h, as determined by LC/MS. Importantly the Ddae linker was stable under these conditions. The crude mixtures were then treated with 1 M hydrazine, pH 7.5 for 3 h to cleave the Ddae linker, remove the helping hand, and generate native GroES sequences (Figures S82, S97).

At this stage, the crude reaction mixture reaction could be injected over RP-HPLC to isolate pure GroES. However, we found that this led to a large yield loss. Instead, taking advantage of the clean one-pot process, we found that GroES in the final reaction could be directly refolded by overnight dialysis (3000 MWCO) into 25 mM Tris, 50 mM NaCl, pH 7.5. The folded heptameric GroES was then purified by SEC (Figure S83, S98) and analyzed under denaturing conditions by HPLC and LC/MS (Figure 4d–g). The L- and D-GroES showed clean products with identical retention times on RP-HPLC (Figure 4d–e) and matched MS to recombinant GroES purified in the same way (pooled from SEC, Figure S99) (Figure 4f–h). We obtained isolated yields of 24 and 23% for folded L- and D-GroES, respectively.

Characterization and activity testing of synthetic L- and D-GroES

With the heptameric proteins (recombinant, synthetic L, and synthetic D) in hand, we performed circular dichroism (CD) spectroscopy to confirm the correctly folded structures. CD (Figure 5a) demonstrated that our synthetic L-GroES possessed structure that mimicked the recombinant control, and synthetic D-GroES showed the inverted spectrum expected for a mirror-image protein.

We then performed enzymatic assays to evaluate the chaperone activity of our synthetic GroES proteins. Here, we first sought to demonstrate that our synthetic L-GroES was fully active compared to recombinant protein in assisting protein folding when combined with recombinant GroEL. Secondly, we tested if D-GroES was capable of forming an active hetero-chiral complex with recombinant GroEL to assist in protein folding.

We used two quantitative folding assays to evaluate GroES activity. We first tested an SR1 assay to evaluate binding of GroES to GroEL. In this assay, functional GroES binds to a single-ring mutant of GroEL (termed SR1) to form a stable, dead-end complex due to the absence of the trans GroEL ring. When saturated with GroES the ATPase activity of SR1 is completely inhibited. In the absence of GroES, the ATPase of SR1 is uninhibited. In the second assay, we directly measured GroEL/ES-mediated refolding of client protein Malate Dehydrogenase (MDH) by GroEL, using synthetic D-GroES with synthetic and recombinant L-GroES as positive controls.

In both the SR1 and MDH assays, the synthetic L-GroES performed similarly to recombinant GroES—validating the high quality of our synthetic material (Figure 5b,c).

However, counter to our expectations, D-GroES was completely non-functional in both the SR1 and MDH assays (Figure 5b,c). Based on these results, we conclude that hetero-chiral GroEL/ES is not functional in protein folding, and unlike substrate recognition/refolding,⁷⁵ there is an element of stereoselectivity to the interaction between GroEL and GroES.

CONCLUSION

In this work, we introduce a general new tool, Fmoc-Ddae-OH, for dealing with poorly soluble and aggregation-prone peptides. Fmoc-Ddae-OH is easy to synthesize and then incorporate into peptide segments for introducing desired solubilizing sequences by Fmoc-SPPS. This new linker can be incorporated at any Lys residue within a peptide segment and is stable to a wide range of conditions allowing handling, purification and storage of the resulting peptides, as well as NCL-based assembly including hydrazide activation and desulfurization. Cleavage can be triggered by mild treatment with aqueous hydrazine in sodium phosphate buffer at near neutral pH. Importantly, the Ddae linker can be cleaved in one pot after NCL and desulfurization, and before a folding step.

We thus conclude that, beyond NCL applications, this new reagent will be generally applicable for installing semi-permanent solubilizing groups in difficult peptides. Moreover, considering the compatibility of its parent Dde protecting group, this strategy is likely also compatible with Boc-SPPS.

We capitalized on these benefits to prepare highly pure forms of both L- and D-GroES heptamers by a NCL strategy that could not be implemented without the help of semi-permanent solubilizing tags. Using these proteins, we show that hetero-chiral L-GroEL/D-GroES is incapable of folding substrate proteins. Thus, although substrate recognition by GroEL/ES is ambidextrous,⁷⁵ the composition of the folding machine is not.

Current work is ongoing to test this new linker in other difficult NCL-based chemical protein synthesis projects and exploit the possibility to screen for a tailor-made solubilizing tag^{86,87} adapted to any difficult segment.⁸⁸ We aim to expand its application beyond the introduction of solubilizing tags, e.g., to immobilize a peptide for solid phase ligations, introduce affinity purification tags, or temporarily link protein subunits in order to facilitate difficult folding processes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We greatly thank Dr Agnès Delmas for fruitful discussions and advice. We thank Rob Marquardt for his mass spectrometry advice and training, and Hervé Meudal, Dr Cyril Colas and the CBM and FR2708 federation NMR and HRMS platforms for analysis of Fmoc-Ddae-OH. Funding for this work was provided by NIH Predoctoral Training Grant F31CA171677 (to M.T.J.), NIH T32 Cardiovascular Training Grant HL007576 (to M.E.P.), NIH Grants AI076168 and AI102347 (to M.S.K.), French National Funding Agency ANR- 15-CE07-022 EasyMiniProt grant (to V.A.) and la Ligue contre le cancer for a postdoctoral fellowship (to M.G.).

ABBREVIATIONS

CD	circular dichroism
Ddae	1-(4,4-dimethyl-2,6-dioxocyclo-hexylidene)-2-[2-(2-aminoethoxy)ethoxy]-propyl
GuHCl	guanidinium chloride
MPAA	4-mercaptophenylacetic acid
NCL	native chemical ligation
SEC	size-exclusion chromatography
SPPS	solid-phase peptide synthesis
TCEP	tris(2-carboxyethyl)phosphine

References

1. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. *Science*. 1994; 266:776. [PubMed: 7973629]
2. Kent SB. *Chemical Society reviews*. 2009; 38:338. [PubMed: 19169452]
3. Hackenberger CP, Schwarzer D. *Angew Chem Int Ed Engl*. 2008; 47:10030. [PubMed: 19072788]
4. Lee CL, Li X. *Curr Opin Chem Biol*. 2014; 22:108. [PubMed: 25299572]
5. Pattabiraman VR, Bode JW. *Nature*. 2011; 480:471. [PubMed: 22193101]
6. Harmand TJ, Murar CE, Bode JW. *Curr Opin Chem Biol*. 2014; 22:115. [PubMed: 25299573]
7. Hemantha HP, Brik A. *Bioorganic & medicinal chemistry*. 2013; 21:3411. [PubMed: 23465437]
8. Spasser L, Brik A. *Angew Chem Int Ed Engl*. 2012; 51:6840. [PubMed: 22696461]
9. Ajish Kumar KS, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A. *Angew Chem Int Ed Engl*. 2009; 48:8090. [PubMed: 19780082]
10. Kumar KS, Spasser L, Erlich LA, Bavikar SN, Brik A. *Angew Chem Int Ed Engl*. 2010; 49:9126. [PubMed: 20815002]
11. Wang P, Dong S, Shieh JH, Peguero E, Hendrickson R, Moore MA, Danishefsky SJ. *Science*. 2013; 342:1357. [PubMed: 24337294]
12. Wang P, Dong S, Brailsford JA, Iyer K, Townsend SD, Zhang Q, Hendrickson RC, Shieh J, Moore MAS, Danishefsky SJ. *Angewandte Chemie International Edition*. 2012; 51:11576.
13. Hirano K, Macmillan D, Tezuka K, Tsuji T, Kajihara Y. *Angewandte Chemie International Edition*. 2009; 48:9557.
14. Murakami M, Okamoto R, Izumi M, Kajihara Y. *Angewandte Chemie International Edition*. 2012; 51:3567.
15. Wilson RM, Dong S, Wang P, Danishefsky SJ. *Angew Chem Int Ed Engl*. 2013; 52:7646. [PubMed: 23775885]
16. Kent SB. *Angew Chem Int Ed Engl*. 2013; 52:11988. [PubMed: 24127351]
17. Chen SY, Cressman S, Mao F, Shao H, Low DW, Beilan HS, Cagle EN, Carnevali M, Gueriguian V, Keogh PJ, Porter H, Stratton SM, Wiedeke MC, Savatski L, Adamson JW, Bozzini CE, Kung A, Kent SB, Bradburne JA, Kochendoerfer GG. *Chemistry & biology*. 2005; 12:371. [PubMed: 15797221]
18. Kochendoerfer GG, Chen SY, Mao F, Cressman S, Traviglia S, Shao H, Hunter CL, Low DW, Cagle EN, Carnevali M, Gueriguian V, Keogh PJ, Porter H, Stratton SM, Wiedeke MC, Wilken J, Tang J, Levy JJ, Miranda LP, Crnogorac MM, Kalbag S, Botti P, Schindler-Horvat J, Savatski L, Adamson JW, Kung A, Kent SB, Bradburne JA. *Science*. 2003; 299:884. [PubMed: 12574628]
19. Milton RC, Milton SC, Kent SB. *Science*. 1992; 256:1445. [PubMed: 1604320]

20. Zhao L, Lu W. *Curr Opin Chem Biol*. 2014; 22:56. [PubMed: 25282524]
21. Zawadzke LE, Berg JM. *Proteins*. 1993; 16:301. [PubMed: 8346193]
22. Verzele D, Maddar A. *Chembiochem*. 2013; 14:1032. [PubMed: 23775826]
23. Raibaut L, El Mahdi O, Melnyk O. *Topics in current chemistry*. 2015; 363:103. [PubMed: 25791484]
24. Zuo C, Tang S, Zheng JS. *Journal of Peptide Science*. 2015; 21:540. [PubMed: 25449562]
25. Shen F, Huang YC, Tang S, Chen YX, Liu L. *Israel Journal of Chemistry*. 2011; 51:940.
26. Olschewski D, Becker CF. *Molecular bioSystems*. 2008; 4:733. [PubMed: 18563247]
27. Saito F, Noda H, Bode JW. *ACS chemical biology*. 2015; 10:1026. [PubMed: 25572124]
28. Malins LR, Payne RJ. *Curr Opin Chem Biol*. 2014; 22:70. [PubMed: 25285753]
29. Johnson EC, Kent SB. *Tetrahedron letters*. 2007; 48:1795. [PubMed: 19177172]
30. Sato T, Saito Y, Aimoto S. *J Pept Sci*. 2005; 11:410. [PubMed: 15635671]
31. Boerema DJ, Tereshko VA, Kent SB. *Biopolymers*. 2008; 90:278. [PubMed: 17610259]
32. Johnson EC, Malito E, Shen Y, Rich D, Tang WJ, Kent SB. *J Am Chem Soc*. 2007; 129:11480. [PubMed: 17705484]
33. Lahiri S, Brehs M, Olschewski D, Becker CF. *Angew Chem Int Ed Engl*. 2011; 50:3988. [PubMed: 21433227]
34. Li X, Kawakami T, Aimoto S. *Tetrahedron letters*. 1998; 39:8669.
35. Wang JX, Fang GM, He Y, Qu DL, Yu M, Hong ZY, Liu L. *Angew Chem Int Ed Engl*. 2015; 54:2194. [PubMed: 25475965]
36. Englebrechtsen DR, Alewood PF. *Tetrahedron letters*. 1996; 37:8431.
37. Hossain MA, Belgi A, Lin F, Zhang S, Shabanpoor F, Chan L, Belyea C, Truong HT, Blair AR, Andrikopoulos S, Tregear GW, Wade JD. *Bioconjug Chem*. 2009; 20:1390. [PubMed: 19552405]
38. Yang SH, Wojnar JM, Harris PW, DeVries AL, Evans CW, Brimble MA. *Organic & biomolecular chemistry*. 2013; 11:4935. [PubMed: 23788006]
39. Harris PW, Brimble MA. *Biopolymers*. 2010; 94:542. [PubMed: 20593475]
40. Peigneur S, Paolini-Bertrand M, Gaertner H, Biass D, Violette A, Stocklin R, Favreau P, Tytgat J, Hartley O. *The Journal of biological chemistry*. 2014; 289:35341. [PubMed: 25352593]
41. Chemuru S, Kodali R, Wetzel R. *Biopolymers*. 2014; 102:206. [PubMed: 24488729]
42. Tan Z, Shang S, Danishefsky SJ. *Proc Natl Acad Sci U S A*. 2011; 108:4297. [PubMed: 21368182]
43. Huang YC, Li YM, Chen Y, Pan M, Li YT, Yu L, Guo QX, Liu L. *Angew Chem Int Ed Engl*. 2013; 52:4858. [PubMed: 23554020]
44. Zheng JS, Yu M, Qi YK, Tang S, Shen F, Wang ZP, Xiao L, Zhang L, Tian CL, Liu L. *J Am Chem Soc*. 2014; 136:3695. [PubMed: 24559202]
45. Johnson T, Quibell M, Sheppard RC. *J Pept Sci*. 1995; 1:11. [PubMed: 9222980]
46. Wahlström K, Planstedt O, Undén A. *Tetrahedron letters*. 2008; 49:3921.
47. Wahlström K, Planstedt O, Undén A. *Tetrahedron letters*. 2008; 49:3779.
48. Zheng JS, He Y, Zuo C, Cai XY, Tang S, Wang ZA, Zhang LH, Tian CL, Liu L. *J Am Chem Soc*. 2016; 138:3553. [PubMed: 26943264]
49. Chhabra SR, Khan AN, Bycroft BW. *Tetrahedron letters*. 1998; 39:3585.
50. Chhabra SR, Hothi B, Evans DJ, White PD, Bycroft BW, Chan WC. *Tetrahedron letters*. 1998; 39:1603.
51. Bycroft BW, Chan WC, Chhabra SR, Hone ND. *Journal of the Chemical Society, Chemical Communications*. 1993:778.
52. Díaz-Mochón JJ, Bialy L, Bradley M. *Organic Letters*. 2004; 6:1127. [PubMed: 15040739]
53. Turner RA, Hauksson NE, Gipe JH, Lokey RS. *Organic Letters*. 2013; 15:5012. [PubMed: 24063667]
54. Chhabra SR, Parekh H, Khan AN, Bycroft BW, Kellam B. *Tetrahedron letters*. 2001; 42:2189.
55. Drinnan N, West ML, Broadhurst M, Kellam B, Toth I. *Tetrahedron letters*. 2001; 42:1159.
56. Chhabra SR, Khan AN, Bycroft BW. *Tetrahedron letters*. 2000; 41:1095.
57. Chhabra SR, Khan AN, Bycroft BW. *Tetrahedron letters*. 2000; 41:1099.

58. Kellam B, Chan WC, Chhabra SR, Bycroft BW. *Tetrahedron letters*. 1997; 38:5391.
59. Yang Y, Verhelst SHL. *Chemical Communications*. 2013; 49:5366. [PubMed: 23648945]
60. Galibert M, Piller V, Piller F, Aucagne V, Delmas AF. *Chemical Science*. 2015; 6:3617.
61. Tan Z, Shang S, Halkina T, Yuan Y, Danishefsky SJ. *J Am Chem Soc*. 2009; 131:5424. [PubMed: 19334683]
62. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR. *Journal of the American Chemical Society*. 1999; 121:11684.
63. Weissenhorn W, Carfi A, Lee KH, Skehel JJ, Wiley DC. *Mol Cell*. 1998; 2:605. [PubMed: 9844633]
64. Malashkevich VN, Schneider BJ, McNally ML, Milhollen MA, Pang JX, Kim PS. *Proc Natl Acad Sci U S A*. 1999; 96:2662. [PubMed: 10077567]
65. Bang D, Kent SB. *Angew Chem Int Ed Engl*. 2004; 43:2534. [PubMed: 15127445]
66. Dang B, Kubota T, Correa AM, Bezanilla F, Kent SB. *Angew Chem Int Ed Engl*. 2014; 53:8970. [PubMed: 24989851]
67. Villain M, Vizzavona J, Rose K. *Chemistry & biology*. 2001; 8:673. [PubMed: 11451668]
68. Arnold RJ, Reilly JP. *Analytical Biochemistry*. 1999; 269:105. [PubMed: 10094780]
69. Plunkett G 3rd, Burland V, Daniels DL, Blattner FR. *Nucleic acids research*. 1993; 21:3391. [PubMed: 8346018]
70. Zheng JS, Tang S, Guo Y, Chang HN, Liu L. *Chembiochem*. 2012; 13:542. [PubMed: 22302623]
71. Fang GM, Wang JX, Liu L. *Angew Chem Int Ed Engl*. 2012; 51:10347. [PubMed: 22968928]
72. Fang GM, Li YM, Shen F, Huang YC, Li JB, Lin Y, Cui HK, Liu L. *Angew Chem Int Ed Engl*. 2011; 50:7645. [PubMed: 21648030]
73. Zheng JS, Tang S, Qi YK, Wang ZP, Liu L. *Nat Protoc*. 2013; 8:2483. [PubMed: 24232250]
74. Siman P, Karthikeyan SV, Nikolov M, Fischle W, Brik A. *Angewandte Chemie International Edition*. 2013; 52:8059.
75. Weinstock MT, Jacobsen MT, Kay MS. *Proc Natl Acad Sci U S A*. 2014; 111:11679. [PubMed: 25071217]
76. Fei X, Ye X, LaRonde NA, Lorimer GH. *Proc Natl Acad Sci U S A*. 2014; 111:12775. [PubMed: 25136110]
77. Nojima T, Ikegami T, Taguchi H, Yoshida M. *Journal of molecular biology*. 2012; 422:291. [PubMed: 22634549]
78. Wang Z, Feng H, Landry SJ, Maxwell J, Gierasch LM. *Biochemistry*. 1999; 38:12537. [PubMed: 10504222]
79. Mascagni P, Tonolo M, Ball H, Lim M, Ellis RJ, Coates A. *FEBS Lett*. 1991; 286:201. [PubMed: 1677897]
80. He QQ, Fang GM, Liu L. *Chinese Chemical Letters*. 2013; 24:265.
81. Wong CTT, Tung CL, Li X. *Molecular bioSystems*. 2013; 9:826. [PubMed: 23302767]
82. Haack T, Mutter M. *Tetrahedron letters*. 1992; 33:1589.
83. Wöhr T, Wahl F, Nefzi A, Rohwedder B, Sato T, Sun X, Mutter M. *Journal of the American Chemical Society*. 1996; 118:9218.
84. Yan LZ, Dawson PE. *J Am Chem Soc*. 2001; 123:526. [PubMed: 11456564]
85. Wan Q, Danishefsky SJ. *Angew Chem Int Ed Engl*. 2007; 46:9248. [PubMed: 18046687]
86. Paradis-Bas M, Albert-Soriano M, Tulla-Puche J, Albericio F. *Organic & biomolecular chemistry*. 2014; 12:7194. [PubMed: 25162452]
87. Schlapschy M, Binder U, Borger C, Theobald I, Wachinger K, Kisling S, Haller D, Skerra A. *Protein engineering, design & selection: PEDS*. 2013; 26:489.
88. Paradis-Bas M, Tulla-Puche J, Albericio F. *Chem Soc Rev*. 2016; 45:631. [PubMed: 26612670]

Figure 1. Fmoc-Ddae-OH linker. (a) Key properties of the new linker. (b) Steps for installing a temporary solubilizing sequence (*helping hand*) using the new linker.

Figure 2.

Incorporation and removal of Lys₆-Ddae in C20 peptide. (a) C20 target amino acid sequence, X indicates position of linker incorporation at Lys residue. (b–f) HPLC traces and deconvoluted MS for C20(Dde) (b), C20 (c), C20(Fmoc-Ddae) (d), C20(Lys₆-Ddae) (e), and purified C20 after linker cleavage (f). (g) Time course of linker cleavage using 1 M hydrazine.

Figure 3.

Synthesis of *E. coli* L31 protein. (a) L31 target amino acid sequence. (b) Amino acid sequences of peptide segments; X indicates position of linker incorporation at Lys residue. Pseudoprolines and (Dmb)Gly are indicated in bold and underline. (c) Synthetic strategy for L31. (d) HPLC traces and deconvoluted MS for initial peptide segments L31-1 (blue), L31-2 (red), and L31-3 (green). (e) HPLC trace and deconvoluted MS of first purified ligation produce L31-1-2. (f) HPLC trace and deconvoluted MS of final, full-length target, L31-1-2-3.

Figure 4. Synthesis of *E. coli* GroES protein. (a) GroES target amino acid sequence. (b) Amino acid sequences of peptide segments; X indicates position of linker incorporation at Lys residue. Pseudoproline and (Dmb)Gly are indicated in bold and underline. (c) Synthetic strategy for GroES. (d) HPLC trace of purified, folded synthetic L-GroES (red). (e) HPLC trace of purified, folded synthetic D-GroES (green). (f–h) Deconvoluted MS for folded synthetic L- (red), D- (green), and recombinant GroES (black), respectively.

Figure 5. Structure and activity of synthetic GroES. (a) Circular dichroism spectra of synthetic L- (red), D- (green), and recombinant (black) GroES. (b) SR1 activity assay. (c) MDH refolding assay. Arrow in (b) shows that additional recombinant GroES can still bind to SR1 in presence of D-GroES.