ORGANIC LETTERS

2009
Vol. 11, No. 18
4144−4147

# Synthesis of N-Terminally Linked Protein Dimers and Trimers by a Combined Native Chemical Ligation-CuAAC Click Chemistry Strategy

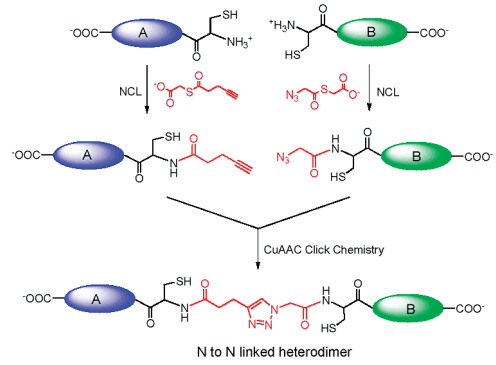
**Junpeng Xiao[†] and Thomas J. Tolbert[*,†,‡]**

*Interdisciplinary Biochemistry Graduate Program and Department of Chemistry, Indiana University, Bloomington, Indiana 47405*

*tolbert@indiana.edu*

**Received July 18, 2009**

## ABSTRACT




**A novel method for the synthesis of N-terminally linked protein multimers is reported. Azide and alkyne thioesters were synthesized for the N-terminal modification of expressed proteins using native chemical ligation (NCL). Proteins modified by these moieties can be joined together to form homodimers and homotrimers via Cu(I)-catalyzed azide−alkyne [3 + 2] cycloaddition (CuAAC) click chemistry. The orthogonal nature of this reaction allows the production of protein heteromultimers, and this is demonstrated by synthesis of a protein heterodimer.**

Multivalent interactions occur in many biological processes such as signal transduction, immune responses, and viral invasion.[1] These multivalent interactions have been the inspiration for the design of peptide and protein multimers for use in biochemical and medicinal applications because of the higher affinity and specificity of multimers compared to monomers for cellular and viral targets. Multivalency has been utilized to improve detection of antigen-specific T-cells by peptide−MHC complexes,[2−4] to produce and improve antiviral peptides against HIV,[5−8] and in vaccine design.[9] Because of the many applications of protein and peptide multimers there is significant interest in their synthesis.

[†] Interdisciplinary Biochemistry Graduate Program.
[‡] Department of Chemistry.

(1) Mammen, M.; Choi, S.-K.; Whitesides, G. M. *Angew. Chem., Int. Ed.* **1998**, *37*, 2754.
(2) Altman, J. D.; Moss, P. A.; Goulder, P. J.; Barouch, D. H.; McHeyzer-Williams, M. G.; Bell, J. I.; McMichael, A. J.; Davis, M. M. *Science* **1996**, *274*, 94.
(3) Crawford, F.; Kozono, H.; White, J.; Marrack, P.; Kappler, J. *Immunity* **1998**, *8*, 675.
(4) Liu, C. P.; Jiang, K.; Wu, C. H.; Lee, W. H.; Lin, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14596.
(5) Li, H.; Song, H.; Heredia, A.; Le, N.; Redfield, R.; Lewis, G. K.; Wang, L. X. *Bioconjugate Chem.* **2004**, *15*, 783.
(6) Li, H.; Guan, Y.; Szczepanska, A.; Moreno-Vargas, A. J.; Carmona, A. T.; Robina, I.; Lewis, G. K.; Wang, L. X. *Bioorg. Med. Chem.* **2007**, *15*, 4220.
(7) Tam, J. P.; Yu, Q. *Org. Lett.* **2002**, *4*, 4167.
(8) Sadler, K.; Zhang, Y.; Xu, J.; Yu, Q.; Tam, J. P. *Biopolymers* **2008**, *90*, 320.

10.1021/ol9016468 CCC: $40.75 
**Published on Web 08/25/2009**

Several approaches have been taken to the synthesis of multivalent peptides and proteins. The most common approach is to utilize amino acid side chains during or after peptide synthesis to create branched structures. This was first applied by Tam to produce multiple antigen peptides (MAPs) by initiating the synthesis of dendrimeric peptides on-resin utilizing both the $N^{\alpha}$ and $N^{\epsilon}$ amino groups of lysine.[9] This approach has been extended by Tam and others to use cysteine and non-natural amino acids to form disulfide-,[10,11] thioether-,[5,6,12,13] thiazolidine-,[7,14] triazole-,[15,16] and oxime-linked[17] multivalent peptides. While these approaches work well for producing multimeric synthetic peptides, the application of these strategies to expressed proteins is complicated by the increased number of amino acid residues in proteins and the necessity of introducing non-natural chemical functionalities onto proteins. To address these issues, the chemoselective reaction expressed protein ligation (EPL) has been utilized to incorporate synthetic peptides and non-natural functionalities onto the C-termini of expressed proteins.[18−21] This has allowed the production of C-terminally linked multimers of expressed proteins.[22,23] Because many proteins cannot be conjugated through their C-termini without affecting biological activity and because it is also desirable to produce protein heteromultimers, we have explored other chemoselective strategies for the production of protein multimers. Herein we report a novel strategy for the synthesis of N-terminally linked homo- and heteromultimers of expressed proteins using a combination of NCL[24,25] with CuAAC click chemistry.[26] This strategy takes advantage of NCL to selectively modify the N-terminus of expressed proteins with azide and alkyne moieties and the bio-orthogonal nature of the CuAAC reaction to produce homo- and heteromultimers.

In our strategy, the target protein is expressed as a fusion protein which contains a tobacco etch virus NIa (TEV) protease cleavable His-tag at the N-terminus. Treatment of the fusion protein with TEV protease produces a protein containing a N-terminal cysteine.[27] Next, azide/alkyne-containing thioesters are specifically ligated to protein N-terminal cysteines via NCL (Figure 1). Finally, the

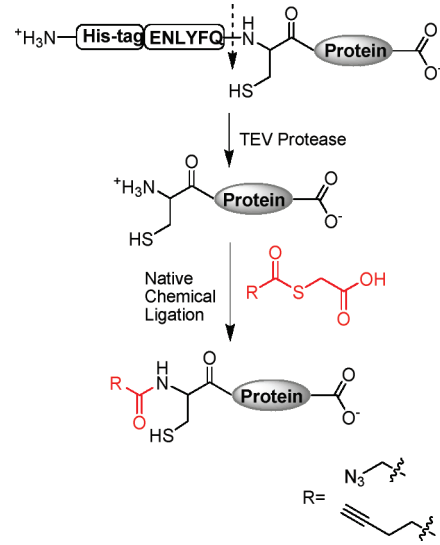



**Figure 1.** Synthesis of azide/alkyne-functionalized proteins.

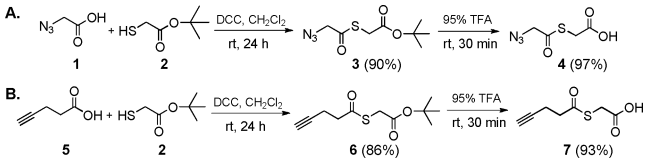
N-terminal azide- and alkyne-containing proteins are either directly coupled together to form dimers or coupled onto alkyne- and azide-containing linkers to form linker-conjugated multimers through CuAAC click chemistry. Using this strategy, not only can N-terminally linked expressed protein homodimers and homotrimers be produced, but synthetically challenging N-terminally linked expressed protein heterodimers can also be produced. In addition to protein multimer formation, expressed proteins that are N-terminally modified by azide and alkyne moieties by the methods described here can also be utilized in the myriad of applications that have been developed for CuAAC click chemistry, such as protein modification, immobilization, and protein−polymer conjugation.[28−36]

Synthesis of an azide-containing thioester was accomplished by coupling azidoacetic acid **1** with *tert*-butyl mercaptoacetate **2** to form *tert*-butyl-protected azidoacetic


(9) Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5409.
(10) Yang, R.; Yang, J.; Weliky, D. P. *Biochemistry* **2003**, *42*, 3527.
(11) Yang, R.; Prorok, M.; Castellino, F. J.; Weliky, D. P. *J. Am. Chem. Soc.* **2004**, *126*, 14722.
(12) Wang, L. X.; Ni, J.; Singh, S. *Bioorg. Med. Chem.* **2003**, *11*, 159.
(13) Zhou, M.; Bentley, D.; Ghosh, I. *J. Am. Chem. Soc.* **2004**, *126*, 734.
(14) Rao, C.; Tam, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 6975.
(15) Franke, R.; Doll, C.; Eichler, J. *Tetrahedron Lett.* **2005**, *46*, 4479.
(16) Galibert, M.; Dumy, P.; Boturyn, D. *Angew. Chem., Int. Ed.* **2009**, *48*, 2576.
(17) Mitchell, J. P.; Roberts, K. D.; Langley, J.; Koentgen, F.; Lambert, J. N. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2785.
(18) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705.
(19) Tolbert, T. J.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *122*, 5421.
(20) Nilsson, B. L.; Hondal, R. J.; Soellner, M. B.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 5268.
(21) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249.
(22) Ziaco, B.; Pensato, S.; D'Andrea, L. D.; Benedetti, E.; Romanelli, A. *Org. Lett.* **2008**, *10*, 1955.
(23) van Baal, I.; Malda, H.; Synowsky, S. A.; van Dongen, J. L.; Hackeng, T. M.; Merkx, M.; Meijer, E. W. *Angew. Chem., Int. Ed.* **2005**, *44*, 5052.
(24) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776.
(25) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705.
(26) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596.
(27) Tolbert, T. J.; Wong, C. H. *Angew. Chem., Int. Ed.* **2002**, *41*, 2171.
(28) Gauchet, C.; Labadie, G. R.; Poulter, C. D. *J. Am. Chem. Soc.* **2006**, *128*, 9274.
(29) Lin, P. C.; Ueng, S. H.; Tseng, M. C.; Ko, J. L.; Huang, K. T.; Yu, S. C.; Adak, A. K.; Chen, Y. J.; Lin, C. C. *Angew. Chem., Int. Ed.* **2006**, *45*, 4286.
(30) Humenik, M.; Huang, Y.; Wang, Y.; Sprinzl, M. *ChemBioChem* **2007**, *8*, 1103.
(31) Moses, J. E.; Moorhouse, A. D. *Chem. Soc. Rev.* **2007**, *36*, 1249.
(32) van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. *Nature* **2007**, *446*, 1105.
(33) Connor, R. E.; Piatkov, K.; Varshavsky, A.; Tirrell, D. A. *ChemBioChem* **2008**, *9*, 366.
(34) Fournier, D.; Du Prez, F. *Macromolecules* **2008**, *41*, 4622.
(35) Gauthier, M. A.; Klok, H. A. *Chem. Commun.* **2008**, 2591.
(36) Merkel, L.; Beckmann, H. S.; Wittmann, V.; Budisa, N. *ChemBioChem* **2008**, *9*, 1220.

acid thioester **3**. The *tert*-butyl-protected **3** was then treated with 95% TFA to give water-soluble azidoacetic acid thioester **4** with an overall 87% yield (Scheme 1A). The same synthetic scheme was also used to synthesize the water-soluble pentynoic acid thioester **7** with an overall 80% yield (Scheme 1B).

**Scheme 1.** Synthesis of Azide- and Alkyne-Containing Thioesters




To demonstrate the feasibility of this strategy, the C2 domain of protein G, an immunoglobulin-binding protein found in group G streptococci,[37] was chosen as a model protein to work with. It was expressed as a fusion protein with a TEV protease cleavable His-tag in *E. coli*.[38] TEV protease cleavage of the protein G fusion protein produced protein G with an N-terminal cysteine (Figure S1A, Supporting Information). The cleaved protein G was separated from the His-tag by $Ni^{2+}$−NTA affinity chromatography and then dialyzed against 20 mM sodium phosphate buffer pH 7.5 at 4 °C for 12 h.

The azide-functionalized protein G **8** was obtained by ligating azidoacetic acid thioester **4** to the N-terminal cysteine of protein G using NCL. The ligation was initiated by adding 4 mM **4** and 30 mM sodium 2-mercaptoethanesulfonate into the protein G solution (4 mg/mL). After 24 h incubation, the reaction mixture was purified by reversed-phase HPLC and lyophilized to give azide−protein G **8**, as confirmed by ESI-MS (Figure S1B, Supporting Information). The ligation of pentynoic acid thioester **7** with cleaved protein G was conducted using the same conditions that were used with **4**. The ligation of **7** proceeded well and produced alkyne-functionalized protein G **9**, as confirmed by ESI-MS (alkyne−protein G **9**, Figure S1C, Supporting Information).

A general procedure for the CuAAC reaction was utilized for the synthesis of protein dimers and trimers. Proteins were dissolved in 1 mL of 50 mM of sodium phosphate buffer pH 7.5, and then $CuSO_4$ and L-ascorbic acid were added to final concentrations of 1 and 2 mM, respectively. Protein multimerization reactions were incubated at room temperature for 12 h and then purified. The N-terminally linked homodimer of protein G **10** was obtained by directly coupling azide−protein G **8** (3 mg/mL, 0.44 mM) with alkyne−protein G **9** (4 mg/mL, 0.59 mM) using the general CuAAC reaction conditions described above (Figure 2A). Purification of the homodimer **10** by reversed-phase HPLC was not possible because **10** has the same retention time as **8** and **9**. Because

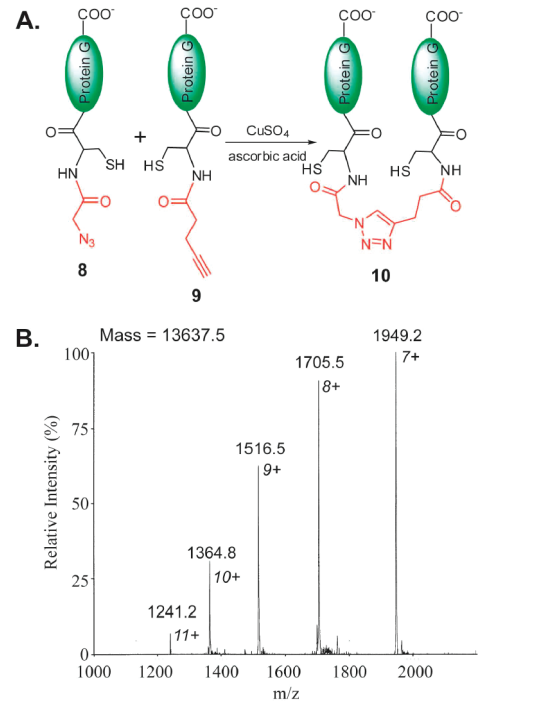



**Figure 2.** (A) Synthesis of protein G homodimer **10**; (B) ESI-MS of **10**, calcd 13637.6, found 13637.5.

of this, size-exclusion chromatography was used to separate **10** from unreacted **8** and **9**. After purification and lyophilization, 5.2 mg of the protein G homodimer **10** was obtained and characterized by analytical HPLC and ESI-MS (Figure S9A (Supporting Information), Figure 2B).

In addition to dimer formation by directly coupling azide and alkyne modified proteins, a diazide linker **11** (the synthesis of **11** is described in the Supporting Information)[39,40] and a trialkyne linker **12** were utilized to demonstrate the application of our strategy for linker-conjugated expressed protein homodimer and homotrimer formation.

The linker-conjugated protein G homodimer **13** was synthesized by a one-pot approach (Figure 3A). The diazide linker **11** (0.2 mM) was incubated with 3 equiv of alkyne−protein G **9** (4 mg/mL, 0.59 mM) under CuAAC reaction conditions. The reaction mixture was then purified by size-exclusion chromatography and lyophilized to give 2.0 mg of the linker-conjugated protein G homodimer **13** (Figure S9B (Supporting Information), Figure 3B). The linker-conjugated homotrimer **14** was also obtained by a one-pot incubation of tripropargylamine **12** (0.15 mM) with approximately 5 equiv of azide−protein G **8** (5 mg/mL, 0.73 mM) using the general CuAAC reaction conditions (Figure 3C). Size-exclusion chromatography purification of the reaction mixture afforded 1.9 mg of the protein G homotrimer **14** (Figure S9C (Supporting Information), Figure 3D).

(37) Sauer-Eriksson, A. E.; Kleywegt, G. J.; Uhlen, M.; Jones, T. A. *Structure* **1995**, *3*, 265.
(38) Xiao, J.; Tolbert, T. J. *Biomacromolecules* **2009**, *10*, 1939.
(39) Fujita, K.; Yamamura, H.; Imoto, T. *J. Org. Chem.* **1985**, *50*, 4393.
(40) Bonger, K. M.; van den Berg, R. J.; Heitman, L. H.; AP, I. J.; Oosterom, J.; Timmers, C. M.; Overkleeft, H. S.; van der Marel, G. A. *Bioorg. Med. Chem.* **2007**, *15*, 4841.

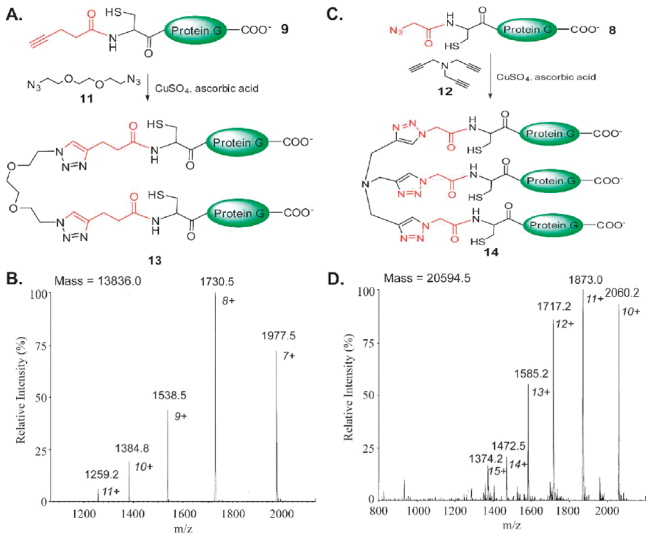



**Figure 3.** (A) Synthesis of linker-conjugated protein G homodimer **13**; (B) ESI-MS of **13**, calcd 13834.6, found 13836.0; (C) synthesis of linker-conjugated protein G homotrimer **14**; (D) ESI-MS of **14**, calcd 20592.1, found 20594.5.

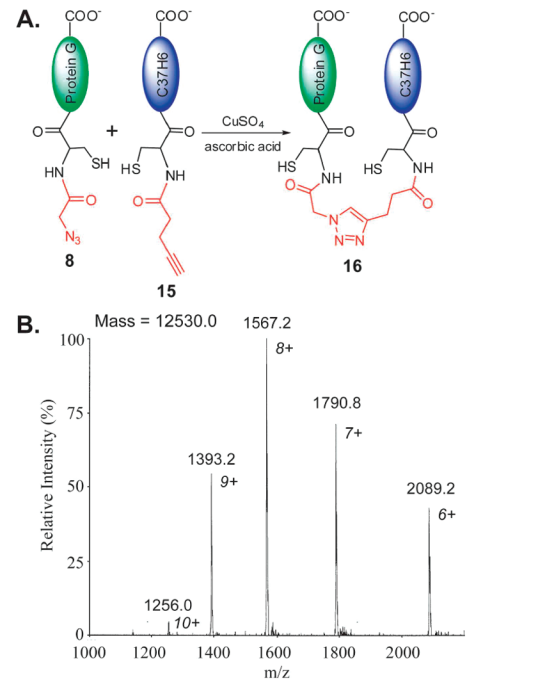



**Figure 4.** (A) Synthesis of protein G−C37H6 heterodimer **16**; (B) ESI-MS of **16**, calcd 12528.3, found 12530.0.

An advantage of using this strategy to produce N-terminally linked protein dimers is that the azide and alkyne moieties can be attached to two different proteins and then be used to produce heterodimers that are synthetically challenging to obtain by other methods. We demonstrate this using azide−protein G **8** and C37H6, an HIV entry inhibitor peptide we have produced by bacterial expression (Figure S2A, Supporting Information).[41] Ligation of **7** to the N-terminal cysteine of C37H6 produced alkyne−C37H6 **15** (Figure S2B, Supporting Information), and then protein G−C37H6 heterodimer **16** was obtained by coupling **8** (3 mg/mL, 0.44 mM) with excess **15** (3.5 mg/mL, 0.61 mM) under the general CuAAC reaction conditions followed by size-exclusion chromatography purification. This afforded 5.0 mg of the heterodimer **16** (Figure S9D (Supporting Information), Figure 4).

In conclusion, a general strategy for the synthesis of N-terminally linked multimeric expressed proteins was developed by combining NCL with CuAAC click chemistry. This strategy utilizes NCL to incorporate the azide and alkyne moieties onto N-terminal cysteines of expressed proteins. The azide- and alkyne-modified proteins can then be coupled together through CuAAC to form N-terminally linked expressed protein homodimers. In addition, linker-conjugated expressed protein homodimers and homotrimers were also obtained in a one-pot reaction between azide- or alkyne-functionalized proteins and alkyne- or azide-containing linkers. This strategy is general and could be extended to other expressed protein multimers, and the linkers could be adjusted by changing the linker length and rigidity. Furthermore, N-terminally linked expressed protein heterodimers, which are difficult to produce selectively by existing methods, can be prepared easily by coupling two different proteins at their N-terminus using the orthogonal CuAAC reaction. Thus, our strategy may be generally applied to prepare multimeric bioactive proteins and peptides for drug development and vaccine design. In addition, modification of proteins with N-terminal azide and alkyne moieties can enable the utilization of other CuAAC click chemistry applications in protein chemistry, such as site-specific protein modification, immobilization, and polymer conjugation.

**Acknowledgment.** This work was supported by Indiana University.

**Supporting Information Available:** Experimental procedures and characterization data. This material is available free of charge via the Internet at http://pubs.acs.org.

OL9016468

(41) Tolbert, T. J.; Franke, D.; Wong, C. H. *Bioorg. Med. Chem.* **2005**, *13*, 909.