NATIVE CHEMICAL LIGATION

Reported by: Yuan Fu   December 7, 2009

INTRODUCTION

Proteins are a class of biological macromolecules that carry out important functions in cells. Because of the importance of proteins, an increasing effort has been devoted to explore their structures and functions. Manipulating the chemical structure of proteins is one commonly used strategy to investigate protein functions. Recombinant expression has enabled the preparation of proteins with various amino acid sequences. However, site specific posttranslational modification and incorporation of unnatural amino acids into proteins remains a challenge. Although unnatural amino acids can be incorporated into proteins by site-directed mutagenesis, the requirement of pre-charged-tRNAs and mutated aminoacyl-tRNA synthetases severely limits the application of this method. Another strategy for protein modification is to chemically synthesize a protein by solid phase peptide synthesis. Once the chemical synthesis of proteins has been achieved, chemists can make any desired change in the structure of the protein molecules. However, it's very difficult to synthesize a peptide that is longer than 50 amino acids due to the accumulation of by-products from incomplete or side reactions and impurities in reagents. To overcome this issue, a convergent strategy has been applied. A set of peptide segments are generated by either solid phase synthesis or recombinant expression and then ligated together with chemical approaches. Chemical approaches for ligation of the two peptide segments include prior thiol capture, native chemical ligation, conformationally assisted ligation and Staudinger ligation, among which native chemical ligation is the most convenient and widely used method.[1] Specifically, native chemical ligation links an N-terminal cysteine-containing peptide to a C-terminal thioester-containing peptide with a native peptide bond at the ligation site (Scheme 1).

DISCOVERY, MECHANISM AND REACTION CONDITION

Native chemical ligation was developed by Kent and Dawson in 1994.[2] The ligation reaction was carried out in buffered aqueous solution with unprotected peptides. In native chemical ligation, the thiol group of the N-terminal cysteine of one peptide attacks the C-terminal thioester of another peptide yielding a thioester-linked intermediate. This intermediate then rearranges irreversibly to form a peptide

bond. (Scheme 1) The first step, which is a transthioesterification reaction, can also occur with cysteine residues imbedded in a peptide sequence, but it does not lead to formation of peptide bond.

Dawson and coworkers have investigated the impact of the C-terminal amino acid of the thioester on the reaction rate. (Scheme 2)[3] Although all 20 amino acids could be applied to the ligation reaction, ligation to peptides terminated by some amino acids, such as proline, valine and isoleucine, occurs slowly.

Scheme 1. The Principle of Native Chemical Ligation
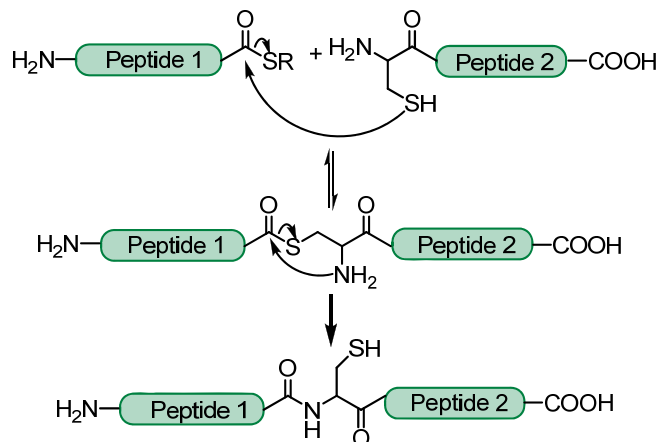
Scheme 2. Influence of C-terminal Amino Acid Residues on Ligation Rate
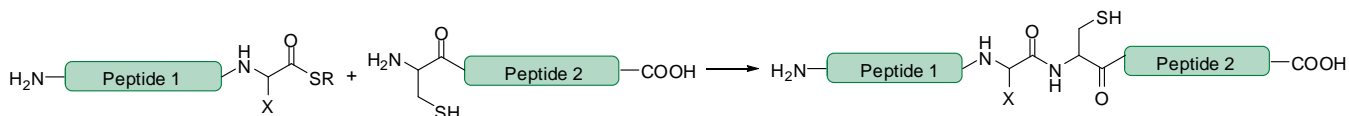
Reaction Rate: X= Gly, Cys, His > Phe, Met, Tyr, Ala, Trp > Asn, Asp, Gln, Glu, Ser, Arg, Lys > Leu, Thr, Val, Ile, Pro

Additional work by Dawson and coworkers has demonstrated that the addition of an aryl thiol such as thiophenol can increase the rate of the native chemical ligation.[4] They explained that aryl thiol additives such as thiophenol convert the alkyl thioester into a more reactive aryl thioester through thiol exchange during the ligation reaction, and this exchange accelerates the native chemical ligation because an aryl thioester is more reactive. (Scheme 3).

Scheme 3. Thiol Exchange between Alkyl and Aryl Thioester
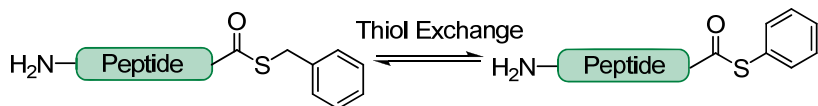
In addition, urea or guanidinium chloride is commonly added in the buffer to prevent aggregation and to enhance the concentration of peptide segments. Detergents, such as dodecylsulfonate (SDS) ,are also used sometimes to improve the solubility of peptide.
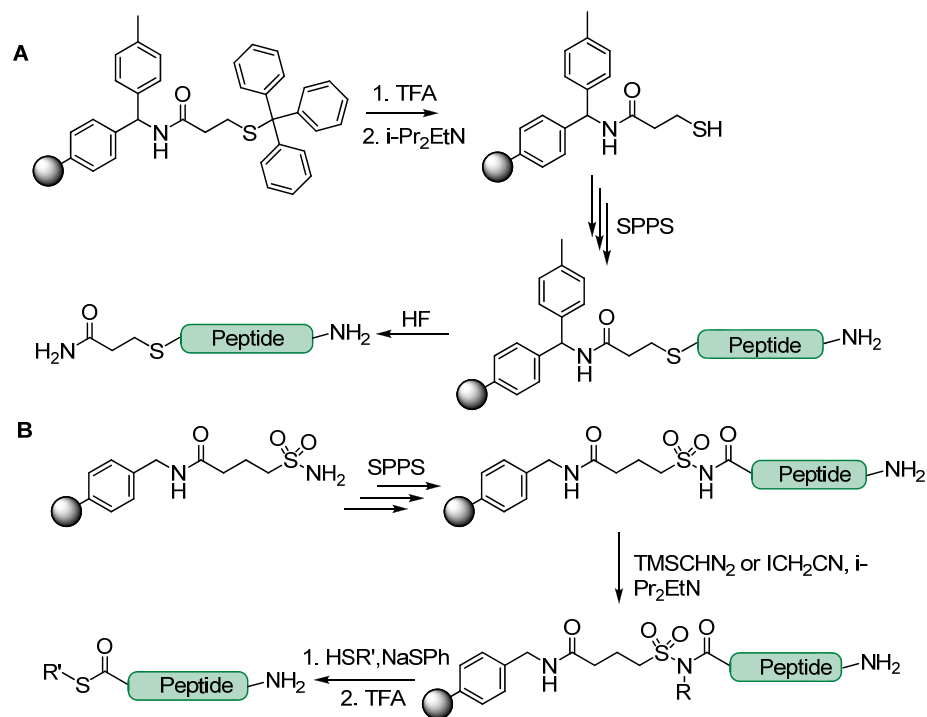
PREPERATION OF STARTING MATERIAL

Both the N-terminal cysteine-containing peptide and C-terminal thioester-containing peptide can be prepared by solid phase peptide synthesis (SPPS) or recombinant protein expression. (Scheme 4) Various

methods have been developed to prepare a peptide thioester and to introduce an N-terminal cysteine into an expressed protein.

Preparation of C-terminal thioester peptides by solid phase peptide synthesis

Scheme 4. Concept of Protein Semisynthesis and Total Synthesis

Zhang and coworkers reported the use 3-mercaptopropionyl MBHA resin for preparation peptide thioester by solid phase peptide synthesis using Boc Chemistry.[5] After extension of the peptide chain on the resin following standard protocols, cleavage by HF yielded the unprotected peptide thioester. (Scheme 5A)
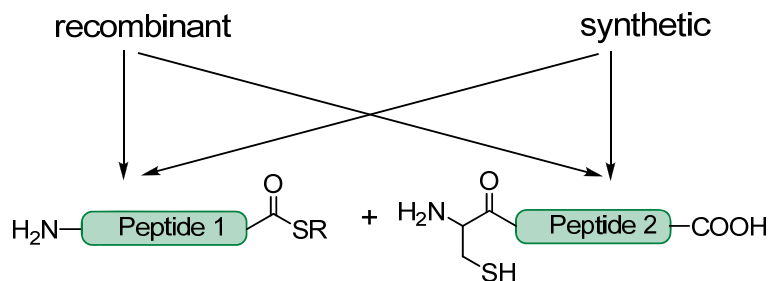
Scheme 5. Generating Peptide Thioester by Solid Phase Synthesis

Alternatively, the peptide thioester can be synthesized by Fmoc solid phase peptide synthesis. Bertozzi and coworkers reported the preparation of sulfamylbutyryl resins for synthesis of peptide thioester and demonstrated its utility in the synthesis of a glycoprotein.[6] After completing the extension of the peptide chain following standard Fmoc chemistry, the sulfonamide was alkylated by iodoacetonitrile and then cleaved from the resin with benzyl mercaptan yielding a peptide thioester. (Scheme 5B)

Recent work by Dawson and coworkers showed that by using benzimidazolinone as a leaving group at the C-terminus of the peptide, a thioester could be generated in situ during the native chemical ligaiton.[7] In this work, the peptide was synthesized on a 3,4-diamino benzamide Rink resin following

standard Fmoc SPPS protocol. After activating the C-terminus through acylation with 4-nitrophenylchloroformate, the desired acylurea peptide was cleaved from the resin by TFA. The resulting peptide could be converted into a thioester or directly used in native chemical ligation.
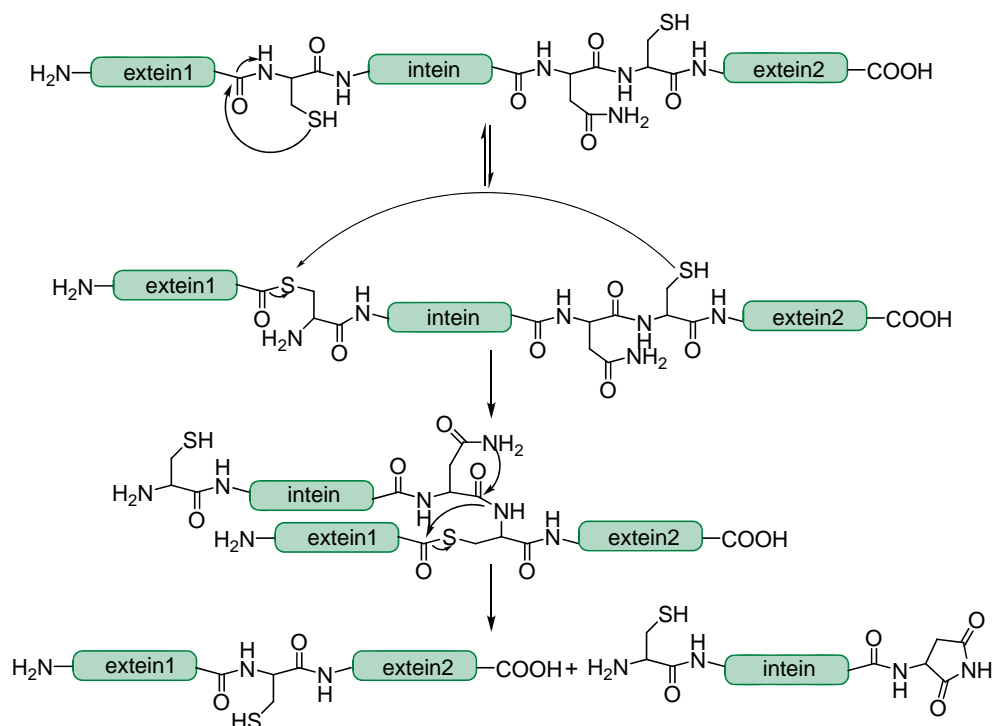
Preparation of C-terminal thioester peptide by expressed protein ligation

Thioesters can also be introduced into a recombinant protein by fusion with an intein. Inteins in protein are analogues of introns in nucleic acids. An intein is a protein segment that excises itself and links the remaining part of protein with a native peptide bond. (Scheme 6) This process, which is called protein splicing, requires no additional enzyme because all enzymatic activities are imbedded in the intein itself. The mechanism of protein splicing is shown in Scheme 6. In protein splicing, a cysteine bound thioester is formed reversibly at the N-terminus of the intein and an intramolecular rearrangement then occurs resulting in the excision of the intein from the protein.

In a particular case, an asparagine to alanine mutation inactivates the splicing but the intein still catalyzes the reversible formation of cysteine-bound thioester during the first step of protein splicing. Xu and coworkers utilized the mutant intein to conjugate a target protein to a chitin binding domain that serves as an affinity tag for purification of a recombinant protein.[8] In their purification system, the target protein was cleaved from the column by transthioesterification, generating a protein thioester. In 1998, Muir and coworkers further developed this method into native chemical ligation for joining a synthetic peptide with a recombinant protein thioester.[9] (Scheme 7)

Scheme 6. Mechanism of Protein Splicing

Scheme 7. Thioesterification of Expressed Protein
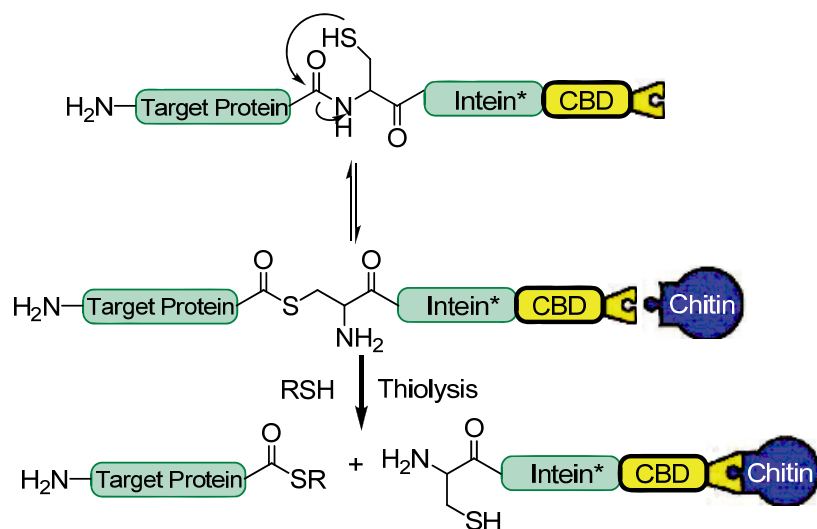
Preparation of N-terminal cysteine peptide by controlled cleavage with protease

Native chemical ligation requires that one peptide fragment has an N-terminal cysteine residue. However, most wildtype proteins start with methionine. Therefore, a system must be created to place a cysteine at 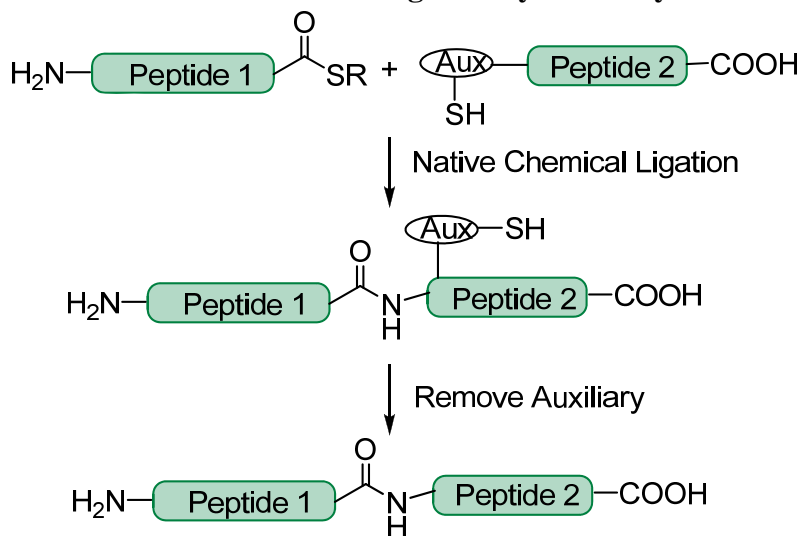the N-terminus of an expressed protein. This is usually achieved by adding a specific cleavage sequence into the fusion proteins and cleaving the protein with a protease. Commonly used proteases include Factor Xa or TEV protease. Muir and coworkers first reported the use of Factor Xa to semisynthesize a segmental isotopically labeled protein for NMR spectroscopy.[10] Wong and coworkers reported the use of TEV protease to produce a glycoprotein.[11] Recently, Leatherbarrow and coworkers also introduced FMDV 3C protease for generating an N-terminal cysteine for native chemical ligation.[12]

LIMITATION OF NATIVE CHEMICAL LIGATION

A severe limitation of native chemical ligation is the requirement of a cysteine at the ligation site. Cysteine is a rare amino acid, comprising only 1.7% of all amino acid residues in proteins. Because it is very difficult to synthesize a peptide that is longer than 50 amino acids, many proteins in nature cannot be obtained through the combination of SPPS and native chemical ligation. Sometimes, an artificial cysteine has to be added into the protein sequence to facilitate the total or semi synthesis of protein. Nevertheless, the artificial cysteine can possibly influence the proper folding of proteins because cysteine is one of the most reactive amino acids and can form disulfide bonds with other cysteine residue. Therefore, some more general ligation methods need to be developed to achieve the cysteine-free native chemical ligation.

One strategy for cysteine-free ligation is to introduce an unnatural thiol containing amino acid surrogate at the ligation site. After native chemical ligation, desulfurization is conducted, and the unnatural amino acid is converted into a natural amino acid. For example, Danishefsky and coworkers reported the use an N-terminal thiol-modified valine derivative reacting with a peptide thioester for native chemical ligation.[13] After ligation, the ligated peptide underwent radical based desulfurization to remove the thiol group. Likewise, in a very recent publication by Liu and coworkers, a thiol group was installed at the N-terminal lysine of a peptide, and one could perform two consecutive ligation steps via the same thiol to synthesize a branched protein.[14] It is worth noting that cysteine imbedded in the peptide sequence should be protected prior to desulfurization in this strategy.

Scheme 8.Native Chemical Ligation by Auxiliary Method

Another similar strategy is to employ a thiol-containing removable auxiliary at the N-terminus of the peptide. The auxiliary acts as a cysteine surrogate and it is removed after ligation (Scheme 8). Botti and Kent have reported the use of the N^{α}-(1-pheynyl-2-mercaptoethyl) as the auxiliary group. [15] TFA or HF was used to remove the auxiliary after native chemical ligation. A drawback for the auxiliary strategy is that the auxililary group will possibly hinder the rearrangement step in native chemical ligation, and the N-terminal amino acid has to be glycine. Alternatively, recent work by Brik and coworkers demonstrated that the auxiliary could be at the side chain of amino acid.[16] In this work, the side chain auxiliary was removed by hydrolysis.

APPLICATION OF NATIVE CHEMICAL LIGATION

Native chemical ligation is particularly useful for the linkage of two peptide segments. This process has been used to synthesize proteins that have been modified by posttranslational modification, unnatural amino acids and biophysical tags. However, the application of native chemical ligation is not

confined to conjugating two peptides. If both an N-terminal cysteine and a C-terminal thioester are present in a same polypeptide, native chemical ligation will result in the formation of a cyclized peptide. Cyclized peptides have improved *in vivo* stability and rigidified structure. The demonstration of this strategy was first reported by Muir and Camarero.[17] In this work, they applied native chemical ligation to cyclize the SH3 domain from the c-Crk adaptor protein.

Moreover, native chemical ligation has been applied to couple dendrimers. It could couple not only two different dendrimers,[18] but also dendrimer with expressed protein.[19] The synthesis of protein dendrimer has been shown in a study by Meijer and Merkx to attach four 27kDa GFP proteins to a 0.7 kDa dendritic core.[19]

Peptide nucleic acids (PNA) are DNA or RNA analogues containing a peudopeptide backbone. Recent work by Burlina and Offer has shown that a PNA thioester could also be used in conjugation with a synthetic peptide by native chemical ligaiton.[20] In principle, it is also possible to conjugate PNA onto a recombinant expressed protein with native chemical ligation, although this has not yet been reported.

In addition to PNA, a protein has been conjugated to a thioester-functionalized DNA by native chemical ligaiton. In a work by Nagamune and coworkers, synthetic DNA was modified at either 5' or 3' and conjugated to an expressed protein with an N-terminal cysteine. [21]

Finally, through native chemical ligation, recombinant expressed proteins were conjugated onto magnetic nanoparticles. For example, CMP-sialic acid synthetase was site-specifically immobilized onto the magnetic nanoparticle by Lin and coworkers.[22] They showed that immobilization on nanoparticles would greatly improve the stability and reusability of the enzyme.

CONCLUSION

Native chemical ligation has proven to be a very useful tool for ligation of peptide segments. Precise alternation of the protein complex can be achieved with the combination of solid phase peptide synthesis, recombinant protein expression, and native chemical ligation. Besides the synthesis of proteins, native chemical ligation has been used for protein conjugation with other macromolecules. Although the requirement of N-terminal cysteine and C-terminal thioester remains a limitation, much effort has been made to generalize native chemical ligation.

For future development, efforts are still needed to develop a simplified and faster cysteine-free ligation method. So far, it is still a challenge to ligate a thioester with an expressed noncysteine terminated protein. The auxiliary strategy could be a possible solution, but methods should be developed to site-specifically label the N-terminal amino acid, especially the N-terminal methionine, of an expressed protein with an auxiliary group. Moreover, the application of native chemical ligation is currently not applicable to *in vivo* work. The reaction condition requires more optimization to reduce toxicity for *in vivo* application.

REFERENCES

(1) Hackenberger, C. P.; Schwarzer, D. *Angew Chem Int Ed* **2008**, *47*, 10030.

(2) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776.

(3) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc Natl Acad Sci U S A* **1999**, *96*, 10068.

(4) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *J Am Chem Soc* **1997**, *119*, 4325.

(5) Zhang, L. S.; Tam, J. P. *J Am Chem Soc* **1999**, *121*, 3311.

(6) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J Am Chem Soc* **1999**, *121*, 11684.

(7) Blanco-Canosa, J. B.; Dawson, P. E. *Angew Chem Int Ed* **2008**, *47*, 6851.

(8) Chong, S.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Kucera, R. B.; Hirvonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M. Q. *Gene* **1997**, *192*, 271.

(9) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc Natl Acad Sci USA* **1998**, *95*, 6705.

(10) Romanelli, A.; Shekhtman, A.; Cowburn, D.; Muir, T. W. *Proc Natl Acad Sci U S A* **2004**, *101*, 6397.

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

(12) Busch, G. K.; Tate, E. W.; Gaffney, P. R. J.; Rosivatz, E.; Woscholski, R.; Leatherbarrow, R. J. *Chem Commun* **2008**, 3369.

(13) Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J. *Angew Chem Int Ed Engl* **2008**, *47*, 8521.

(14) Yang, R. L.; Pasunooti, K. K.; Li, F. P.; Liu, X. W.; Liu, C. F. *J Am Chem Soc* **2009**, *131*, 13592.

(15) Botti, P.; Carrasco, M. R.; Kent, S. B. H. *Tetrahedron Lett* **2001**, *42*, 1831.

(16) Lutsky, M. Y.; Nepomniaschiy, N.; Brik, A. *Chem Commun* **2008**, 1229.

(17) Camarero, J. A.; Muir, T. W. *J Am Chem Soc* **1999**, *121*, 5597.

(18) Dirksen, A.; Meijer, E. W.; Adriaens, W.; Hackeng, T. M. *Chem Commun* **2006**, 1667.

(19) van Baal, I.; Malda, H.; Synowsky, S. A.; van Dongen, J. L. J.; Hackeng, T. M.; Merkx, M.; Meijer, E. W. *Angew Chem Int Ed* **2005**, *44*, 5052.

(20) Burlina, F.; Dixson, D. D.; Doyle, R. P.; Chassaing, G.; Boddy, C. N.; Dawson, P.; Offer, J. *Chem Commun* **2008**, 2785.

(21) Takeda, S.; Tsukiji, S.; Ueda, H.; Nagamune, T. *Org Biomol Chem* **2008**, *6*, 2187.

(22) Yu, C. C.; Lin, P. C.; Lin, C. C. *Chem Commun* **2008**, 1308.