Available online at www.sciencedirect.com
SCIENCE DIRECT®
Tetrahedron Letters 47 (2006) 5219–5223
Tetrahedron Letters

Reiterative cysteine-based coupling leading to complex, homogeneous glycopeptides

Bin Wu,[a] J. David Warren,[a] Jiehao Chen,[a] Gong Chen,[a] Zihao Hua[a] and Samuel J. Danishefsky[a,b,*]

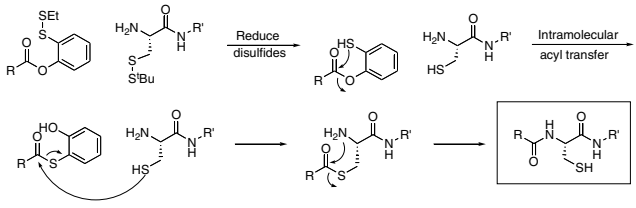
[a]*Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10021, USA*
[b]*Department of Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, USA*

Received 28 March 2006; revised 3 April 2006; accepted 5 April 2006
Available online 5 June 2006

Abstract—Reiterative approaches in the fashioning of erythropoietin-directed, polyglycosylated polypeptides are disclosed.

Our laboratory is engaged in an ongoing program devoted to the development of methodologies for the de novo synthesis of homogeneous glycoproteins. Glycoproteins constitute an important class of biomacromolecules, and a great deal of effort has been directed toward the understanding of the role of glycosylation in various critical protein functions, such as protein folding, proteolytic stability, and cell adhesion.[1] Furthermore, a number of glycoproteins are known to possess valuable and exploitable therapeutic activities. Prominent examples, which have served to focus our endeavors in this field, include erythropoietin,[2] which is commonly used in the treatment of anemia, the various isoforms of prostate specific antigen,[3] and candidate antigens which could serve as the basis for gp120-directed HIV vaccines.[4] Presumably, Nature's complex apparatus for accomplishing posttranslational glycosidation is there to impart some advantage to the glycoprotein product. However, despite considerable interest, the field of glycobiology faces a nontrivial obstacle to the rigorous investigation of the implications of protein glycosidation. The isolation of significant quantities of homogeneous glycoprotein from natural sources is exceedingly difficult.[5] It is our vision that, through recourse to the medium of total synthesis, it might be possible to prepare, de novo, meaningful amounts of structurally homogeneous glycoproteins including, in time, erythropoietin.

Our entry into the arena of glycoprotein synthesis arose as a consequence of a long-standing involvement in the field of oligosaccharides as tumor and HIV antigens. In pursuit of the more complex goal of fully synthetic complex glycoproteins, we have adopted a long-term, phased approach commencing with the development of enabling methodologies. Thus, we first learned how to construct N-linked glycopeptides by preparing, through total synthesis, complex oligosaccharide units and appending them to small peptide fragments through a Kochetkov–Lansbury amination–aspartylation protocol.[6] We next developed a means to ligate a fully synthetic carbohydrate–peptide domain to another peptide fragment using cysteine-based native chemical ligation.[7] In order to achieve the level of convergence necessary to prepare *multiply* glycosylated proteins, it would be critical to develop a means by which to couple two fully functionalized carbohydrate–peptide fragments. Toward this end, we recently disclosed a new paradigm which accomplishes in situ generation of a thioester from a glycopeptide possessing a C-terminal phenolic ester (Scheme 1). This fragment can then be coupled with



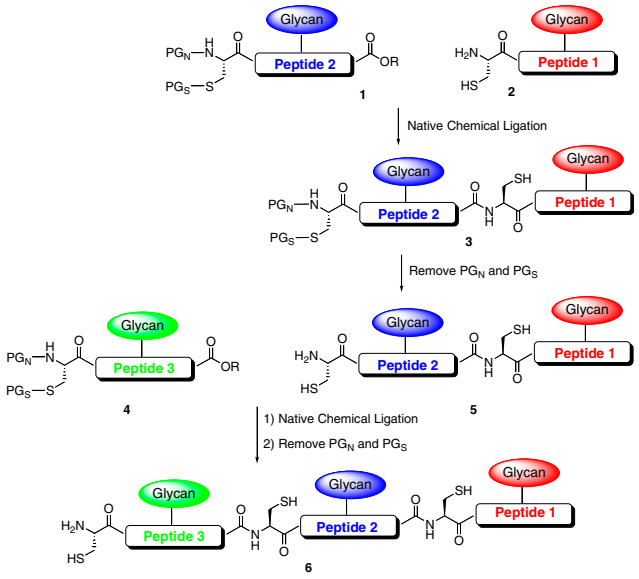
Scheme 1. Cysteine-based native chemical ligation of glycopeptides.

* Corresponding author. Tel.: +1 212 639 5501; fax: +1 212 772 8691; e-mail: danishes@mskcc.org

0040-4039/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetlet.2006.04.132

a second glycopeptide displaying an N-terminal cysteine residue to provide a homogeneous, fully synthetic polypeptide incorporating two N-linked oligosaccharide domains.[8]

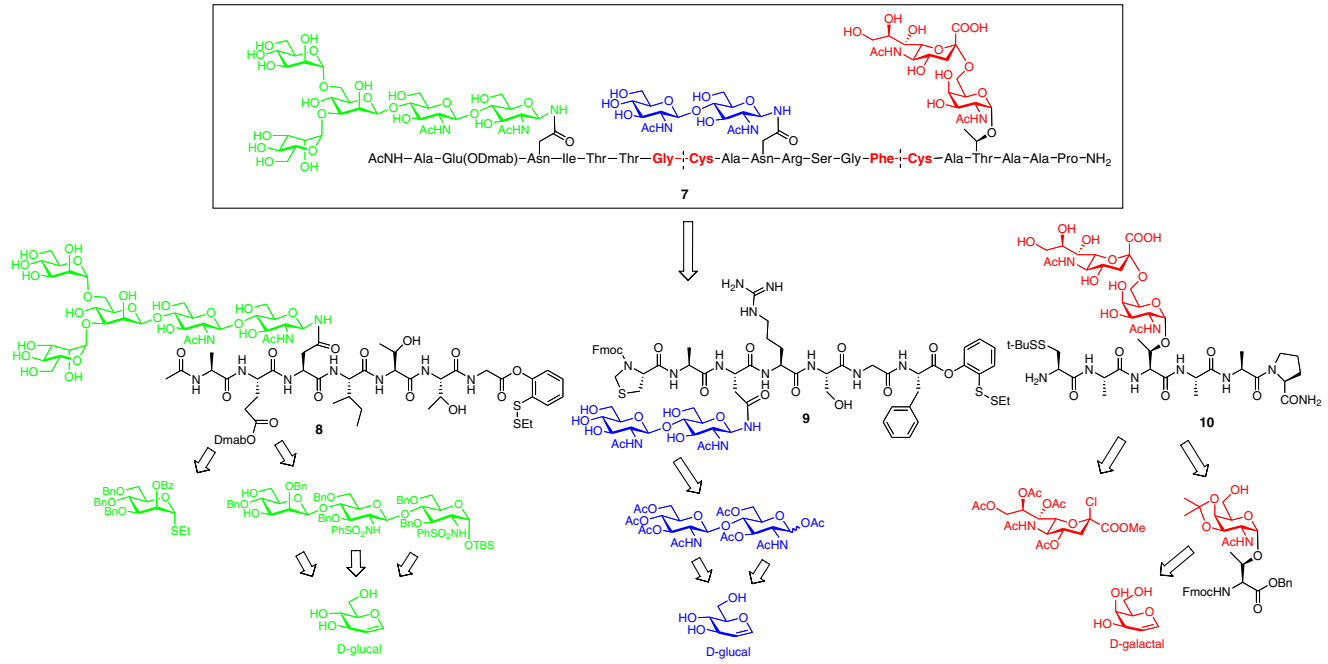
To further explore the outer limits of this type of transformation, as well as its applicability to the generation of complex glycoproteins displaying *multiple* sites of glycosylation, we required the capacity for reiterative coupling, even in the context of highly complex settings. Thus, as outlined in Scheme 2, two glycopeptide fragments (**1** and **2**) would be joined in the first step through expansion on the native chemical ligation method. The C-terminal coupling partner (**1**) would possess a masked cysteine residue at its N-terminus. Following removal of the N- and S- protecting groups, the newly formed glycopeptide (**5**) would be coupled once again with a third synthetic glycopeptide unit (**4**). Deprotection of the N-terminus and the cysteine residue of the coupled product would provide a large polypeptide unit containing three different oligosaccharide domains (**6**). In considering a protecting group strategy, we came to favor a 1,3-thiazolidine-4-carboxo (Thz) group, which Kent and coworkers had employed in their total synthesis of crambin.[9] We projected that this group would be cleavable under mild conditions which are compatible with survival of the potentially labile glycopeptide ensemble. There would thus be exposed a new N-terminal cysteine-based acyl acceptor for reiteration of the scheme and elongation of the ensemble.



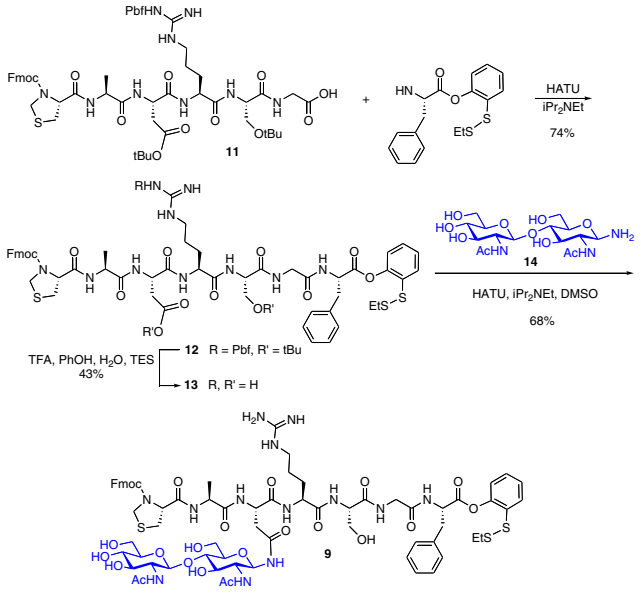
Scheme 2. General strategy for reiterative coupling of glycopeptides.

As a proof of concept, we set as our target the glycopeptide **7**, comprised of 20 amino acids and three different oligosaccharide domains, two of which contain N-linkages to Asn residues, and one of which is O-linked to a Thr residue (Scheme 3). This is the sort of capability that must be well in hand if a venture to be directed to erythropoietin is to be feasible. According to our plan, this complex structure would be assembled through reiterative couplings of the individual glycopeptide fragments (cf. **8–10**). The coupling precursor fragments leading to the **8–10** domains were synthesized according to established precedents, drawing frequently from the logic of glycal assembly honed over many years in our laboratory.[10]

Peptide **11** was prepared through Fmoc based synthesis (Scheme 4). Coupling of **11** with the phenolic ester of phenylalanine, as shown, provided peptide **12**, possessing the appropriate functionalization at the future C-terminal coupling site. Following removal of the amino



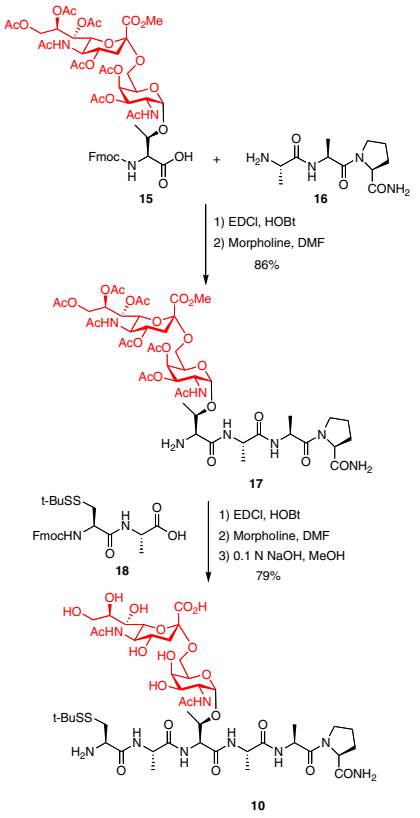
Scheme 3. Synthetic strategy for multiply glycosylated peptide **7**.



Scheme 4. Preparation of coupling partner **9**.

acid protecting groups (cf. **12** to **13**), aspartylation[11] with glycosylamine **14** (prepared from chitobiose through amination[11a]) provided **9** in 68% yield.

The preparation of the O-linked glycopeptide, **10**, is outlined in Scheme 5. As shown, the synthesis commenced with the previously described glycosylamino acid, **15**.[12] The remainder of the peptide was then appended to the N- and C-termini of the threonine glycosylamino acid to afford, following deprotection, the coupling partner **10**, possessing an N-terminal cysteine residue.



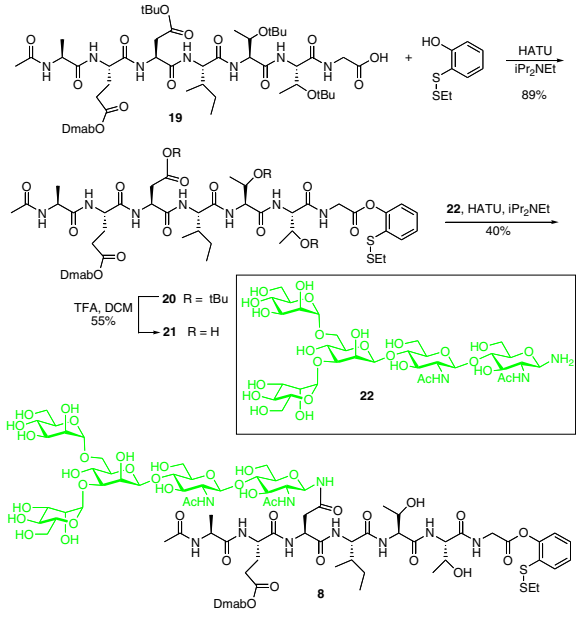
Scheme 5. Preparation of coupling partner **10**.

As shown in Scheme 6, peptide **19**, prepared through solid phase synthesis, was advanced to **8** by recourse to standard esterification, deprotection, and aspartylation techniques. We note that, in the case of **8**, the glutamic acid was equipped with the readily removable Dmab protecting group, thereby enabling selective appendage of the oligosaccharide domain to the aspartic acid.[13]

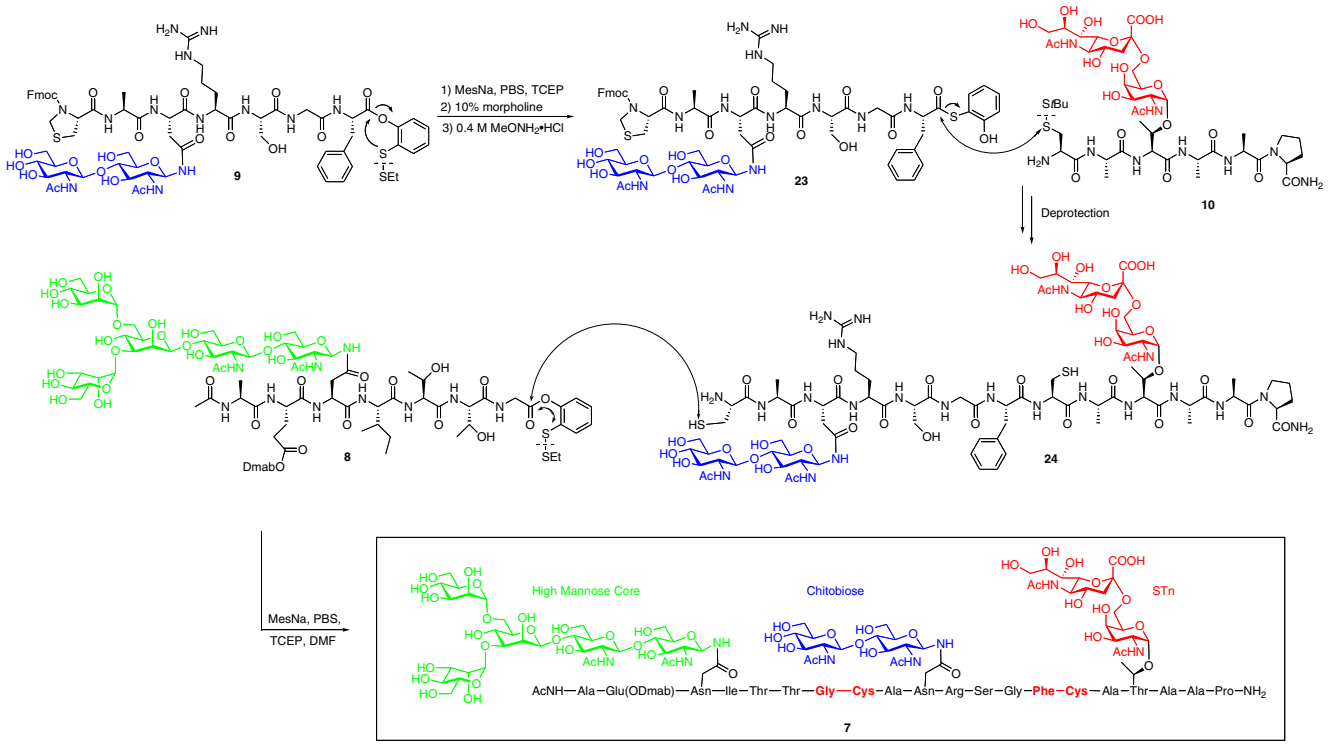
With the key building blocks in hand, we were prepared to evaluate the feasibility of the first stage of our reiterative coupling protocol. In the event, upon treatment with MesNa (2-mercaptoethanesulfonic acid sodium salt) in a phosphate buffered saline solution (pH = 7.4), fragments **9** and **10** smoothly underwent ligation within 20 h, as indicated by LC–MS analysis, to afford a bifunctional glycopeptide in approximately 50% yield (Scheme 7). Happily, the required acyl migration, transferring the entire C-terminal glycopeptidyl domain from O to S (cf. structure **23**) had indeed occurred. The latter is apparently converted to the MesNa ester, which subsequently undergoes NCL with the N-terminal cysteine residue of **10**.

The next challenge would be that of unmasking the N-terminal Thz and Fmoc functionalities in preparation for the second coupling reaction. This turned out to be a nontrivial matter. However, optimal deprotection conditions were ultimately developed, in which the glycopeptide was treated with 10% morpholine in DMF, followed by an aqueous solution of $MeONH_2 \cdot HCl$, to afford **24** in 57% yield.

Finally, the second native chemical ligation event, between **24** and **8** proceeded cleanly under the previously



Scheme 6. Preparation of coupling partner **8**.



Scheme 7. Synthesis of **7** through reiterative glycopeptide coupling sequence.

described conditions to provide the complex, structurally homogeneous glycopeptide **7**, for the moment in 38% yield.

In summary, we have demonstrated the synthesis of a multifunctional glycopeptide through reiterative native chemical ligation. The compatibility of both N-linked and O-linked glycans in this process is noteworthy. Its ability to encompass the biologically critical sialic acid glycosides is particularly encouraging. In our ongoing quest to develop methodologies to enable the total synthesis of multiply glycosylated complex glycoproteins of potential clinical value, this new reiterative coupling protocol constitutes an important, if early, entry. We expect that the strategies and protocols of the type disclosed herein will be extendable in our quest to build homogeneous complex glycoproteins of value. An important hurdle to be overcome in this pursuit required solutions to the problem of noncysteine-based ligation. Dramatic progress in this regard will be disclosed soon.

Acknowledgements

This work was supported by the NIH (CA28824 to S.J.D.). We thank Dr. George Sukenick, Ms. Sylvi Rusli and Ms. Hui Fang of the Sloan-Kettering Institute's NMR core facility for mass spectral and NMR spectroscopic analysis (SKI core Grant No.: CA02848). Postdoctoral fellowship support is gratefully acknowledged by BW (New York State Department of Health, New York State Breast Cancer Research and Education Fund) and JDW (NIH, CA62948).

References and notes

1. (a) Varki, A. *Glycobiology* **1993**, *3*, 97–130; (b) Helenius, A. *Mol. Biol. Cell* **1994**, *2*, 253–265; (c) Riederer, M. A.; Hinnen, A. *J. Bacteriol.* **1991**, *173*, 3539–3546; (d) Klausner, R. D.; Sitia, R. *Cell* **1990**, *62*, 611–614; (e) Roth, J. *Chem. Rev.* **2002**, *102*, 285–303; (f) Grogan, M. J.; Pratt, M. R.; Marcaurelle, L. A.; Bertozzi, C. R. *Annu. Rev. Biochem.* **2002**, *71*, 593–634; (g) Nilsson, B. L.; Soellner, M. B.; Raines, R. T. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91–118; (h) Ratner, D. M.; Adams, E. W.; Disney, M. D.; Seeberger, P. H. *ChemBioChem.* **2004**, *5*, 1375–1383.
2. (a) Szymkowski, D. E. *Curr. Opin. Drug Discov. Devel.* **2005**, *8*, 590–600; (b) Pavlou, A. K.; Reichert, J. M. *Nat. Biotechnol.* **2004**, *22*, 1513–1519; (c) Jelkmann, W.; Wagner, K. *Ann. Hematol.* **2004**, *83*, 673–686; (d) Ridley, D. M.; Dawkins, F.; Perlin, E. *J. Natl. Med. Assoc.* **1994**, *86*, 129–135.
3. (a) Okada, T.; Sato, Y.; Kobayashi, N.; Sumida, K.; Satomura, S.; Matsuura, S.; Takasaki, M.; Endo, T. *Biochim. Biophys. Acta* **2001**, *1525*, 149–160; (b) Dudkin, V. Y.; Miller, J. S.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 736–738.
4. (a) Geyer, H.; Holschbach, C.; Hunsmann, G.; Schneider, J. *J. Biol. Chem.* **1988**, *263*, 11760–11767; (b) Mandal, M.; Dudkin, V. Y.; Geng, X.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2004**, *43*, 2557–2561; (c) Geng, X.; Dudkin, V. Y.; Mandal, M.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2004**, *43*, 2562–2565.
5. Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631–664.
6. (a) Wang, Z. G.; Zhang, X.; Visser, M.; Live, D.; Zatorski, A.; Iserloh, U.; Lloyd, K. O.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2001**, *40*, 1728–1732; (b) Wang, Z. G.; Zhang, X.; Live, D.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2000**, *39*, 3652–3656; (c) Wang, Z. G.; Warren, J. D.; Dudkin, V. Y.; Zhang, X.; Iserloh, U.; Visser, M.;

Eckhardt, M.; Seeberger, P. H.; Danishefsky, S. J. *Tetrahedron* **2006**, *62*, 4954–4978.
7. For the key paper teaching the powers of cysteine-based native chemical ligation, see: (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776–779; (b) Miller, J. S.; Dudkin, V. Y.; Lyon, G. J.; Muir, T. W.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2003**, *42*, 431–434.
8. Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 6576–6578.
9. Bang, D.; Kent, S. B. H. *Angew. Chem. Int. Ed.* **2004**, *43*, 2534–2538.
10. Danishefsky, S. J.; Bilodeau, M. T. *Angew. Chem. Int. Ed.* **1996**, *35*, 1380–1419.
11. (a) Likhosherstov, L. M.; Novikova, O. S.; Derevitskaja, V. A.; Kochetkov, N. K. *Carbohydr. Res.* **1986**, *146*, C1–C5; (b) Cohen-Anisfeld, S. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1993**, *115*, 10531–10537.
12. Schwarz, J. B.; Kuduk, S. D.; Chen, X. T.; Sames, D.; Glunz, P. W.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1999**, *121*, 2662–2673.
13. Chan, W. C.; Bycroft, B. W.; Evans, D. J.; White, P. D. *J. Chem. Soc., Chem. Commun.* **1995**, *21*, 2209–2210.