Available online at www.sciencedirect.com
ScienceDirect
Tetrahedron Letters
Tetrahedron Letters 48 (2007) 1795–1799

Towards the total chemical synthesis of integral membrane proteins: a general method for the synthesis of hydrophobic peptide-$^{\alpha}$thioester building blocks

Erik C. B. Johnson[a,b] and Stephen B. H. Kent[a,b,c,*]

[a]*Institute for Biophysical Dynamics, 929 E. 57th Street, Chicago, IL 60637, USA*
[b]*Department of Biochemistry, 929 E. 57th Street, Chicago, IL 60637, USA*
[c]*Department of Chemistry, 929 E. 57th Street, Chicago, IL 60637, USA*

Received 23 August 2006; revised 4 January 2007; accepted 5 January 2007
Available online 10 January 2007

Abstract—Modification of a peptide-$^{\alpha}$thioester with a sequence of six arginines on the thioester leaving group can render soluble all peptides derived from a polytopic integral membrane protein. This strategy greatly simplifies the synthesis of peptide-$^{\alpha}$thioester building blocks for the total chemical synthesis of integral membrane proteins by native chemical ligation.

Integral membrane proteins constitute approximately one-third of all proteins expressed in the cell,[1] yet our knowledge of how they function on a molecular level remains rudimentary. Compared to most soluble, cytosolic proteins, integral membrane proteins present numerous experimental difficulties. One of the foremost challenges is the isolation of adequate amounts of protein for study, as the overexpression of most integral membrane proteins fails to produce protein material in sufficient amounts for biophysical studies. Various strategies that rely on optimizations to recombinant expression techniques have been proposed to address this problem,[2–5] although none has proven general. An alternative to recombinant overexpression for the production of integral membrane proteins may be total chemical protein synthesis. Chemical protein synthesis has been used to prepare a large number of soluble proteins for study,[6] and offers unique atom-by-atom control over the covalent structure of the protein that often greatly assists biophysical and functional studies. In addition, once a synthesis has been established, the amount of protein material obtainable depends only upon the scale of synthesis.

Like recombinant overexpression of integral membrane proteins, the total chemical synthesis of integral membrane proteins presents special challenges. The first and foremost of these challenges is the poor solubility of the peptide building blocks used to assemble the protein. Unlike peptides derived from water-soluble globular proteins, some peptides derived from integral membrane proteins contain stretches of amino acids that reside within the lipid bilayer ('transmembrane' (TM) sequences), and are therefore highly hydrophobic. This inherent hydrophobicity complicates peptide workup and purification in the commonly used mixed aqueous/organic solvents. Strategies to improve the handling and purification of hydrophobic peptides have largely focused on the use of various organic co-solvents after cleavage from the resin, and modifications to HPLC mobile phases.[7–12] While these strategies have shown some degree of success in the synthesis of small oligomeric channel proteins,[10,13,14] which contain transmembrane helices that face water and therefore have a polar character, they have not proven general.

Recently, Aimoto suggested the use of an Arg5 tag in the thioester moiety to improve the handling properties of a transmembrane peptide for use with native chemical ligation.[15] We set out to explore the general utility of this approach, using diacylglycerol kinase (DGK)—a 121-residue membrane protein from *Escherichia coli*[15,16]—as a model system. The topology and

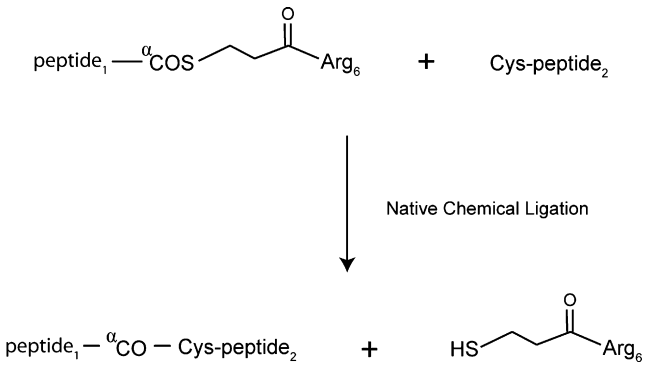
Keywords: Chemical protein synthesis; Integral membrane proteins; Diacylglycerol kinase; Native chemical ligation; Thioester.
* Corresponding author. Tel.: +1 773 702 4912; fax: +1 773 702 0439; e-mail: skent@uchicago.edu

0040-4039/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetlet.2007.01.030

secondary structure of DGK, based on prediction and experimental evidence[17–19] is shown in Figure 1. The protein is composed of three transmembrane helices and an amphipathic helical region, with the residues critical for catalysis located in the second cytoplasmic loop. DGK has been extensively studied as a model integral membrane protein,[20–22,19] although its three-dimensional structure has not been solved experimentally. It was reasoned that the methods developed to synthesize DGK would prove general to many other integral membrane proteins. DGK also provides a convenient assay to eventually determine the identity of the final folded synthetic protein, as it catalyzes the phosphorylation of diacylglycerol and other lipids for the purposes of lipid bilayer maintenance.

Our initial synthetic designs for DGK consisted of ligation of either two ~60 residue unprotected peptides or three ~40 residue unprotected peptides.[23] Both of these strategies resulted in peptides that were extremely insoluble in most solvents after cleavage from the polymer support, as well as refractory to analysis by analytical reversed-phase HPLC (no elution from the column) and both electrospray and MALDI mass spectrometry (poor or no ionization).

Rather than modifying standard procedures for peptide workup, analysis and purification of soluble peptides, we decided to modify the hydrophobic peptide as shown in Scheme 1 in order to be compatible with standard protocols for peptide and protein synthesis. In chemical protein synthesis by native chemical ligation, an N-terminal cysteine peptide reacts with a C-terminal peptide thioester to generate a native amide bond at the ligation site.[23] During this reaction, the thiol component of the thioester is the leaving group, and does not remain in the final ligated product. Therefore, in order to solubilize the peptide during workup and purification by HPLC, we took advantage of the constitutively charged, highly soluble guanidinium group by adding arginine residues to the thioester leaving group. After ligation, these residues would no longer remain part of the product peptide, and thus the native sequence would be recovered. Based on previous work with peptides containing single TM domains,[24,25] six arginine residues were added to the thioester leaving group.



Scheme 1. Charge tag strategy. A hydrophobic peptide ($peptide_1$) is synthesized to include a sequence of six arginine residues on the thioester moiety. During ligation to a Cys-peptide, the arginine modifications are removed as part of the thioester leaving group.

The revised synthetic strategy for DGK consisted of five peptide-$^{\alpha}$thioesters encompassing residues 1–112, each of which carries an 'Arg tag' on the thioester moiety. Residues 113–121 would be synthesized on a water-compatible cross-linked polymer as a handle for performing solid-phase chemical ligation as we recently reported,[26] since otherwise the ligated product polypeptide would not be expected to be soluble in the aqueous ligation buffers after loss of the arginine residues (recently reported water-soluble thiol catalysts[27] could also be modified with an Arg tag to maintain peptide solubility during thioester activation). Ligation sites were chosen to give peptides of <40 amino acids in length and to be compatible with proper folding and activity of the enzyme.[20] Except for the peptide containing residues 1–25 of the protein, all other four peptides have a net positive hydropathy score based on the Kyte–Doolittle hydropathy index,[28] and can therefore be classified as 'hydro-

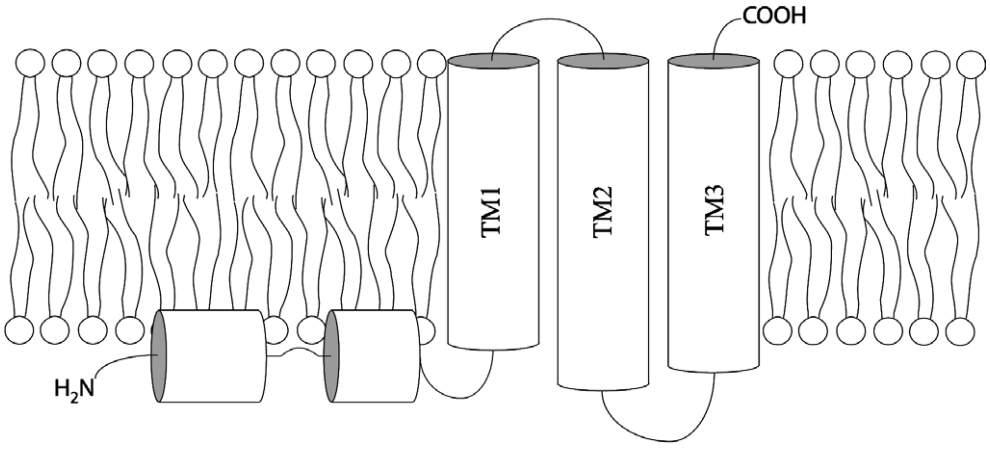


Figure 1. Topology and secondary structure of DGK. The N-terminal region of the protein is composed of two amphipathic helices, followed by three transmembrane helices connected by a very short extracellular loop and an intracellular loop that contains part of the catalytic apparatus of the protein. Depending on the actual length of the TM segments and the degree of membrane burial of the amphipathic helices, between 50% and 70% of the protein resides in the lipid bilayer.

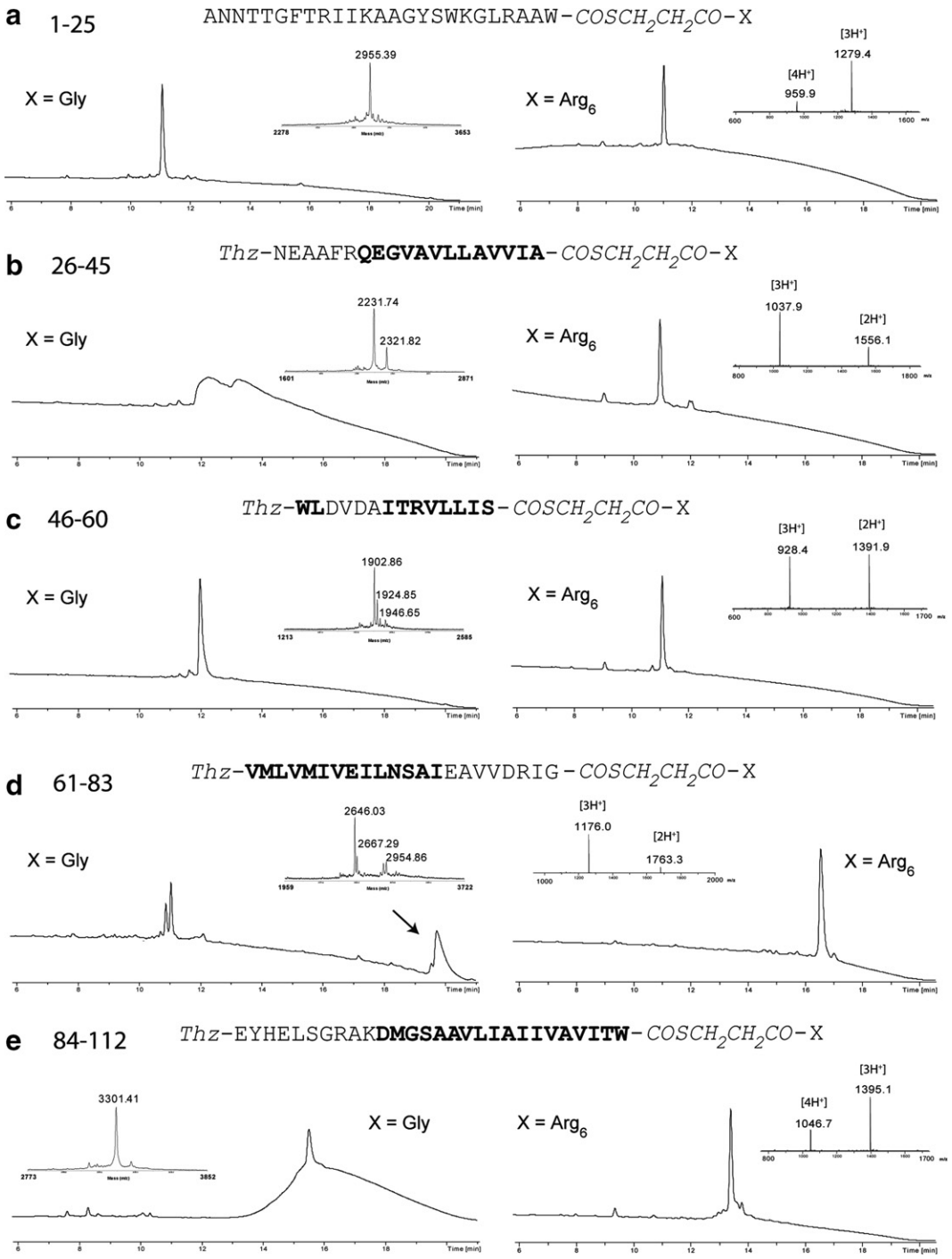


Figure 2. HPLC and mass analysis of crude DGK peptide building blocks with and without an arginine tag. Residues thought to reside within the lipid bilayer are in bold. (a) Peptide DGK (1–25). No significant difference is seen in the quality of synthesis or retention behavior on RP-HPLC with an arginine tag (X = Gly, expected mass 2954.3 Da, observed mass 2954.4 ± 1.5 Da; X = Arg6, expected mass 3835.4 Da, observed mass 3835.4 ± 0.3 Da). (b) Peptide DGK (26–45). Addition of an arginine tag significantly improves elution behavior on RP-HPLC (X = Gly, expected mass 2229.6 Da, observed mass 2230.7 ± 1.1 Da, the +90 peak is likely a *p*-cresol ester byproduct from HF cleavage;[29] X = Arg6, expected mass 3110.7 Da, observed mass 3110.5 ± 0.4 Da). (c) Peptide DGK (46–60). The arginine tag shifts elution of the peptide to a slightly earlier time (X = Gly, expected mass 1901.2 Da, observed mass 1901.9 ± 1.0 Da, the +22 peaks are sodium adducts; X = Arg6, expected mass 2782.3 Da, observed mass 2782.0 ± 0.3 Da). (d) Peptide DGK (61–83). Inclusion of an arginine tag shifts elution to an earlier time and greatly improves recovery from the reversed-phase column (X = Gly, expected mass 2644.2 Da, observed mass 2645.0 ± 1.3 Da, the +21 peak is likely a sodium adduct, and the larger mass species is likely a byproduct from HF cleavage; X = Arg6, expected mass 3525.3 Da, observed mass 3524.8 ± 0.3 Da). (e) Peptide DGK (84–112). The arginine-tagged peptide shows dramatically improved resolution on RP-HPLC, eluting in a single sharp peak rather than a broad hump (X = Gly, expected mass 3301.8 Da, observed mass 3300.4 ± 1.7 Da; X = Arg6, expected mass 4182.9 Da, observed mass 4182.6 ± 0.4 Da). Thioester is –$COSCH_2CH_2CO$–, and Thz is (4*R*)-1,3-thiazolidine-4-carboxylic acid. Masses for the X = Gly peptides were obtained by MALDI-TOF MS of the crude cleavage mixture, and masses for the X = Arg6 peptides were obtained by ESI-MS over the principal peak during LC–MS analysis.

phobic' based on physiochemical parameters (see Supplementary data).

The analytical results for all five peptides synthesized *without* arginine modification to the thioester leaving group are shown in Figure 2 (left). Peptide 1–25 was soluble in 50% acetonitrile after cleavage from the resin, and eluted as a symmetrical peak on HPLC. Peptide 26–45, which contains a portion of TM1, was initially soluble in the same solvent after cleavage, but formed a gel upon standing. After increasing the amount of TFA to ~60% to redissolve the gel, a sample of the peptide solution was injected onto an RP-HPLC column and eluted under standard gradient conditions. The peptide eluted as an undefined hump, which is often typical of strongly hydrophobic peptides (if elution is observed at all, as noted above). Peptide 46–60 strongly precipitated after workup in 50% acetonitrile, and was finally dissolved by increasing the TFA concentration to 50%. This peptide, while quite insoluble, eluted in a symmetrical peak on HPLC. Peptide 61–83 precipitated on standing, and was dissolved in 50% TFA. This peptide eluted poorly as a broad peak at ~65% acetonitrile. Peptide 84–112, while surprisingly soluble in 50% acetonitrile, eluted as a broad hump on HPLC similar to peptide 26–45. MALDI-TOF mass analysis of all crude peptides showed that the target peptides were the predominant species in each crude mixture (Fig. 2 (left, insets)).

In contrast to these results, introduction of a 'charge tag' on the thioester moiety dramatically improved the handling properties and analytical behaviour of all five peptides. Analytical data for crude products of the same peptides synthesized *with* a six-arginine tag on the thioester moiety are also shown in Figure 2 (right). Peptide 1–25 was highly soluble after cleavage from the resin, and showed similar behavior on HPLC as the untagged peptide. Peptide 26–45 was soluble after cleavage, did not form a gel on standing, and eluted in a sharp peak slightly earlier than the untagged peptide. Peptide 46–60 was highly soluble on workup, and eluted slightly earlier on HPLC than the control peptide. Peptide 61–83 with an arginine tag was soluble and did not precipitate on standing. The addition of the arginine tag sharpened the elution peak and shifted elution to an earlier time by 3 min, or by 12% acetonitrile. Addition of the arginine tag to peptide 84–112 remarkably improved its elution profile on HPLC by sharpening the broad hump into a predominantly single peak and shifting elution to a significantly lower percentage of acetonitrile. Electrospray mass analysis over the main peak by LC–MS showed each crude peptide was the expected product (Fig. 2 (right, insets)).

The use of an Arg tag has greatly simplified the preparation of peptides necessary for the synthesis of membrane proteins. Using this Arg tag approach, we have been able to purify on a preparative scale hundreds of milligrams of homogenous integral membrane peptide building blocks derived from DGK, as well as hydrophobic peptides derived from other integral membrane proteins such as KcsA and signal peptide peptidase (data not shown). The Arg tag approach has also been applied with great effect to many other soluble protein targets that contain poorly soluble peptide building blocks (e.g., the HIV-1 protease, RNaseA and lysozyme), and thus appears to have a general utility in chemical protein synthesis.

We have also tested lysine residues as part of a charge tag, which show a similar effect, although the guanidinium group is preferable because it is more soluble in water.[28] Additional arginines beyond six confer marginal improvements to elution time and solubility (data not shown). The Arg tag could also be used as a handle to synthesize and purify poorly soluble peptides that will not be used for chemical protein synthesis by native chemical ligation.[30,31] After purification, the Arg tag could be removed by hydrolysis of the thioester under basic conditions.[32]

In summary, we have used the Arg tag strategy to prepare large amounts of peptide-$^{\alpha}$thioester building blocks for the total chemical synthesis of an integral membrane protein. We are currently optimizing solid phase ligation methods for the total synthesis of the full-length DGK polypeptide using these peptide-thioester Arg tag building blocks.

Acknowledgements

E.C.B.J. is supported by the MD/Ph.D. Graduate Training in Growth and Development program at the University of Chicago (NIH T32 HD007009). We also gratefully acknowledge support from the USA National Institutes of Health (Grant NIH R01 GM075993-01) and the USA Department of Energy 'Genomes to Life-Genomics' Program (Grant DE-FG02-04ER63786).

Supplementary data

Synthetic strategy for DGK, hydrophobicity of peptide building blocks, materials and experimental procedures. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tetlet.2007.01.030.

References and notes

1. Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029–1038.
2. Loll, P. J. *J. Struct. Biol.* **2003**, *142*, 144–153.
3. Sun, J.; Savva, C. G.; Deaton, J.; Ronald Kaback, H.; Svrakic, M.; Young, R.; Holzenburg, A. *Arch. Biochem. Biophys.* **2005**, *434*, 352–357.
4. Grisshammer, R.; Tate, C. G. *Q. Rev. Biophys.* **1995**, *28*, 315–422.
5. Grisshammer, R.; White, J. F.; Trinh, L. B.; Shiloach, J. *J. Struct. Funct. Genomics* **2005**, *6*, 159–163.
6. Dawson, P. E.; Kent, S. B. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
7. Fisher, L. E.; Engelman, D. M. *Anal. Biochem.* **2001**, *293*, 102–108.
8. Lew, S.; London, E. *Anal. Biochem.* **1997**, *251*, 113–116.

9. Bollhagen, R.; Schmiedberger, M.; Grell, E. *J. Chromatogr. A* **1995**, *711*, 181–186.
10. Kochendoerfer, G. G.; Salom, D.; Lear, J. D.; Wilk-Orescan, R.; Kent, S. B.; DeGrado, W. F. *Biochemistry* **1999**, *38*, 11905–11913.
11. Kassim, S. Y.; Restrepo, I. M.; Kalivretenos, A. G. *J. Chromatogr. A* **1998**, *816*, 11–20.
12. Glover, K. J.; Martini, P. M.; Vold, R. R.; Komives, E. A. *Anal. Biochem.* **1999**, *272*, 270–274.
13. Clayton, D.; Shapovalov, G.; Maurer, J. A.; Dougherty, D. A.; Lester, H. A.; Kochendoerfer, G. G. *PNAS* **2004**, *101*, 4764–4769.
14. Kochendoerfer, G. G.; Jones, D. H.; Lee, S.; Oblatt-Montal, M.; Opella, S. J.; Montal, M. *J. Am. Chem. Soc.* **2004**, *126*, 2439–2446.
15. Schneider, E. G.; Kennedy, E. P. *Biochim. Biophys. Acta* **1976**, *441*, 201–212.
16. Loomis, C. R.; Walsh, J. P.; Bell, R. M. *J. Biol. Chem.* **1985**, *260*, 4091–4097.
17. Sanders, C. R.; Czerski, L.; Vinogradova, O.; Badola, P.; Song, D.; Smith, S. O. *Biochemistry* **1996**, *35*, 8610–8618.
18. Oxenoid, K.; Sonnichsen, F. D.; Sanders, C. R. *Biochemistry* **2002**, *41*, 12876–12882.
19. Oxenoid, K.; Kim, H. J.; Jacob, J.; Sonnichsen, F. D.; Sanders, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 5048–5049.
20. Wen, J.; Chen, X.; Bowie, J. U. *Nature Struct. Biol.* **1996**, *3*, 141–148.
21. Badola, P.; Sanders, C. R. *J. Biol. Chem.* **1997**, *272*, 24176–24182.
22. Zhou, Y.; Bowie, J. U. *J. Biol. Chem.* **2000**, *275*, 6975–6979.
23. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776–779.
24. Melnyk, R. A.; Partridge, A. W.; Yip, J.; Wu, Y.; Goto, N. K.; Deber, C. M. *Biopolymers* **2003**, *71*, 675–685.
25. Sato, T.; Saito, Y.; Aimoto, S. *J. Pept. Sci.* **2004**.
26. Johnson, E. C.; Durek, T.; Kent, S. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 3283–3287.
27. Johnson, E. C.; Kent, S. B. *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.
28. Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105–132.
29. Miranda, L. P.; Jones, A.; Meutermans, W. D. F.; Alewood, P. F. *J. Am. Chem. Soc.* **1998**, *120*, 1410–1420.
30. Englebretsen, D. R.; Alewood, P. F. *Tetrahedron Lett.* **1996**, *37*, 8431–8434.
31. Becker, C. F.; Oblatt-Montal, M.; Kochendoerfer, G. G.; Montal, M. *J. Biol. Chem.* **2004**, *279*, 17483–17489.
32. Camarero, J. A.; Adeva, A.; Muir, T. W. *Lett. Pept. Sci.* **2000**, *7*, 17–21.