Research Article

Journal of PeptideScience

Received: 29 November 2011 Revised: 4 January 2012 Accepted: 10 January 2012 Published online in Wiley Online Library: 19 March 2012

(wileyonlinelibrary.com) DOI 10.1002/psc.2401

Native chemical ligation in dimethylformamide can be performed chemoselectively without racemization

Marc Dittmann,[a] Muheeb Sadek,[a,b] Ralf Seidel[a] and Martin Engelhard[a]*

Native chemical ligation of unprotected peptides in organic solvents has been previously reported as a fast, efficient, and suitable method for coupling of hydrophobic peptides. However, it has not been determined whether the reaction can be carried out without possible side reactions or racemization. Here, we present a study on the chemoselectivity of this method by model reactions designed to test the reactivity of Arg and Lys side chains as well as that of α-amino groups. A possible racemization of the C-terminal amino acid of the N-terminal peptide was also investigated. The results show that ligation in organic solvents can be conducted chemoselectively without side reactions with other nucleophilic groups. Furthermore, no racemization of the C-terminal amino acid was observed if both educts were added simultaneously. Thus, native chemical ligation can be performed either in aqueous buffer systems or in organic solvents paving the way for the synthesis of larger hydrophobic peptides and/or membrane proteins. Copyright © 2012 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: native chemical ligation in organic solvents; peptide synthesis; hydrophobic peptides; racemization

Introduction

Native chemical ligation (NCL) is based on a thioester-mediated chemoselective reaction of two unprotected peptide segments resulting in a native peptide bond at the ligation site [1]. To undergo NCL, the N-terminal peptide has to possess a C-terminal thioester, whereas the C-terminal peptide contains an N-terminal Cys. During the past 15 years, NCL has been developed into a routinely used tool for the synthesis of proteins soluble in aqueous buffers [2] that even allowed the synthesis and chemical modification of proteins of up to 304 amino acids [3]. However, ligation of hydrophobic peptides and/or proteins such as membrane proteins is hampered by their low solubility and their tendency to aggregate under standard aqueous conditions [4,5]. Various attempts have been published to overcome these difficulties utilizing aqueous buffer systems. For example, ligations have been carried out in the presence of organic solvents, denaturants such as urea or guanidinium salts, and/or detergents such as DDM, SDS, and DPC [6–11]. However, a more general method is not yet available.

Recently, we have described NCL of model peptides in organic solvents under purely anhydrous conditions (Figure 1) [12]. In this work, it has been shown that DMF readily solubilizes hydrophobic peptides, and in the presence of a base, like TEA as well as thiol additives, a fast and efficient NCL is observed. To improve the solubility of hydrophobic products, different other additives were also tested (e.g. α-cyclodextrin, sodium trifluoroacetate, and lithium chloride [14,15]). Best results regarding solubility and yields were obtained with thiophenol in the presence of LiCl. Under these conditions, the reaction was already completed in less than 3 h.

Although these results demonstrate that NCL can be performed in organic solvents with high yields, the question of possible side reactions has not been addressed so far. For a general application of NCL in organic solvents, it is mandatory that the ligation has to proceed chemoselectively in the presence of other nucleophilic groups. Furthermore, during activation of the α-COOH of the unprotected N-terminal peptide, organic bases might trigger racemization at this site. Here, we demonstrate that NCL in organic solvents is chemoselective and occurs without racemization.

Materials and Methods

Boc-L-Leu-Pam resin was obtained from NeoMPS (Strasbourg, France). Boc-protected amino acids and HBTU were purchased from Merck Chemicals (Nottingham, UK). DCM and DMF were obtained from Applied Biosystems (Darmstadt, Germany). TFA was purchased from Roth (Karlsruhe, Germany). All other chemicals were obtained from Sigma Aldrich (Taufkirchen, Germany) with the highest purity available.

Synthesized peptides were analyzed by RP-HPLC on a Beckman (Krefeld, Germany; System Gold; modules 126, 168

* *Correspondence to: Martin Engelhard, Max Planck Institute of Molecular Physiology, Otto Hahn Str. 11, 44227 Dortmund, Germany. E-mail: martin.engelhard@mpi-dortmund.mpg.de*

a *Max Planck Institute of Molecular Physiology, Otto Hahn Str. 11, 44227 Dortmund, Germany*

b *Present address: Institut für Organische und Biomolekulare Chemie, Georg-August-Universität Göttingen, Tammannstr. 2, 37077 Göttingen, Germany*

Abbreviations used: *DDM, n-dodecyl-β-D-maltoside; DPC, dodecylphosphocholine; HBTU, O-benzotriazole-N,N,N,N-tetramethyl-uronium-hexafluorophosphate; Leu, L-leucyl; MPAA, (4-carboxylmethyl) thiophenol; NCL, native chemical ligation; SLeu, $-SCH_2COLeu$; TEA, triethylamine; TFA, trifluoroacetic acid.*

Journal of PeptideScience

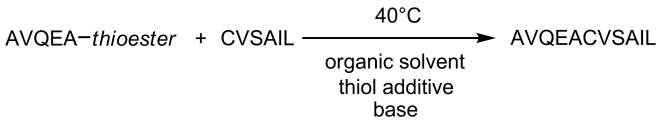

Figure 1. General scheme of NCL reactions in organic solvents (thioester = $-SCH_2CO$*Leu*) [12]. The sequences of the peptides were derived from transmembrane-spanning helices of the archaeal transducer protein *Np*HtrII [13].

and 508) dual-pump high-pressure mixing system, 214 nm UV and 280 nm detection, and a Bischoff C18 column (250 × 4.6 mm, 5 μm). The molecular masses of the peptides were determined by ESI-MS on an ESI-MS Finnigan LCQ Advantage MAX (San Jose, USA) apparatus. The preparative purifications were carried out on a Bischoff C18 column (250 × 20 mm, 5 μm) using a Waters System (Eschborn, Germany; Controller 600, Dual λ Absorbance detector 2487) equipped with a fraction collector.

Peptide Synthesis

All peptides were synthesized manually using Boc-chemistry, *in situ* neutralization, and HBTU activation protocols on a 0.2 mmol synthetic scale [16]. The N-terminal peptides Ala-Val-Gln-Glu-Ala-S*Leu* and Ala-S*Leu* were synthesized on a resin that generates a C-terminal thioester after HF cleavage [17]. The peptides were deprotected and cleaved off from the resin by anhydrous HF and *p*-cresol as scavenger for 1 h at 0 °C. Peptides were purified by HPLC on a preparative C18 column using linear gradients from buffer A (0.1% (v/v) TFA in water) to B (0.08% (v/v) TFA in acetonitrile). Fractions were analyzed by ESI-MS and HPLC on an analytical C18 column. Mass spectrometric data of synthesized peptides are summarized in Table 1.

Ligation Experiments in Organic Solvents

Peptides (~3.5 mM) were dissolved in DMF or DMF in the presence of 0.18 M LiCl. A thiol additive was added in various concentrations (see Results and Discussion section). TEA was added to a final concentration of 20 mM. The reaction mixtures were stirred at 40 °C. At each time point, an aliquot (3 μL) was withdrawn and the reaction quenched by adding 20 μL of 50% acetonitrile in water (0.1% TFA). After reduction of possible S–S bridges by tris(2-carboxyethyl) phosphine, the crude product mixture was analyzed by HPLC on a C18 column (detection wavelength 214 nm). The amount of educts, aryl thioester, and products were determined by integration of peak areas using 32 Karat Software (Beckman Coulter). Because the extinction coefficients of the aryl thioesters were considerably higher than those of the alkyl thioesters, the values for the aryl thioesters were divided by an empirical determined factor of 1.9. To determine the relative amount of each compound in the reaction mixture, the corrected peak areas were divided by the number of peptide bonds. Fractions of reactants were finally calculated by dividing the corrected peak areas by the sum of integrated educt and product areas.

Table 1. Mass spectrometric data of peptides

Peptide	Theoretical $[M+H]^+$	Experimental $[M+H]^+$
Ala-Val-Gln-Glu-Ala-S*Leu*	704.8	704.4
Cys-Val-Ser-Ala-Ile-Leu	605.7	605.5
Ala-Val-Ser-Ala-Ile-Leu	573.2	573.7
Lys-Val-Ser-Ala-Ile-Leu	630.8	630.7
Ala-S*Leu*	277.4	277.3
Cys-Val	221.3	221.1
Ala-Cys-Val	292.4	292.3
D-Ala-Cys-Val	292.4	292.3
Ala-Val-Gln-Glu-Ala-Cys-Val-Ser-Ala-Ile-Leu	1104.3	1104.2
Ala-Val-Gln-Glu-a-Cys-Val-Ser-Ala-Ile-Leu	1104.3	1104.0

Results and Discussion

In the previous work, two reaction conditions were found suitable for NCL in organic solvents, which differ in the thiol additive, namely, thiophenol or (4-carboxylmethyl) thiophenol (MPAA). Because both thiols lead to fast reactions and high yields, these conditions were analyzed for possible side reactions. Amino acids whose side chains might react with activated thioesters are Lys and Arg. Additionally, free N-terminal amino groups could also react. As an example, potential side reactions for Lys during NCL are depicted in Figure 2.

A second serious problem concerns a decrease of the pK_a of the α-CH group adjacent to the C-terminal carboxyl group during its activation that could lead to racemization. In the following text, experimental results are presented addressing these possible side reactions.

Side Chain or N-Terminal Directed Reactions

To test whether nucleophilic amino groups can react directly with C-terminal thioesters, we synthesized two C-terminal peptides lacking an N-terminal Cys. These peptides contained either an N-terminal Ala (Ala-Val-Ser-Ala-Ile-Leu) or an N-terminal Lys (Lys-Val-Ser-Ala-Ile-Leu). The reaction of Ala-Val-Ser-Ala-Ile-Leu or Lys-Val-Ser-Ala-Ile-Leu with Ala-Val-Gln-Glu-Ala-S*Leu* could lead to possible products **3**, **4**, and **5** (Figure 2).

In a first set of experiments, the reactions were carried out in DMF using concentrations of 3.5 mM for each peptide in the presence of 20 mM TEA, 0.3 M thiophenol, and 0.18 M LiCl. After quenching the reaction with 50% acetonitrile in water (0.1% TFA) at various time points, the crude product mixture was analyzed by HPLC on a C18 column. HPLC traces from typically ligation experiments are shown in Figure 3.

The ligation between Ala-Val-Gln-Glu-Ala-S*Leu* with Cys-Val-Ser-Ala-Ile-Leu yields already after 1 min the transesterification product Ala-Val-Gln-Glu-Ala-S*Ph* (Figure 3a; peak at 20.8 min). After 10 min, this reaction is almost completed. Concomitantly, the final product Ala-Val-Gln-Glu-Ala-Cys-Val-Ser-Ala-Ile-Leu (peak at 21.3 min) is formed. On the other hand, by replacing Cys-Val-Ser-Ala-Ile-Leu by Lys-Val-Ser-Ala-Ile-Leu (Figure 3b),

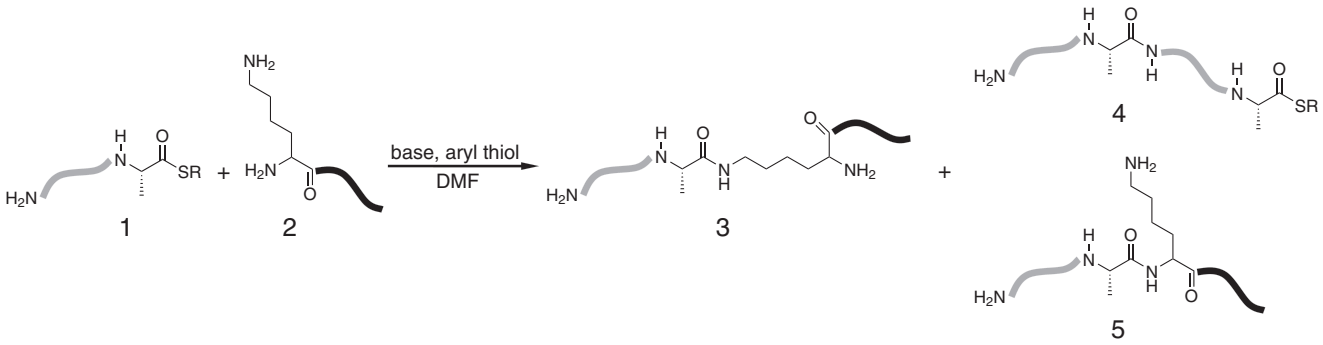

Figure 2. Possible side reactions of NCL in organic solvents. Possible side reactions of the guanidinium group of Arg are not depicted.

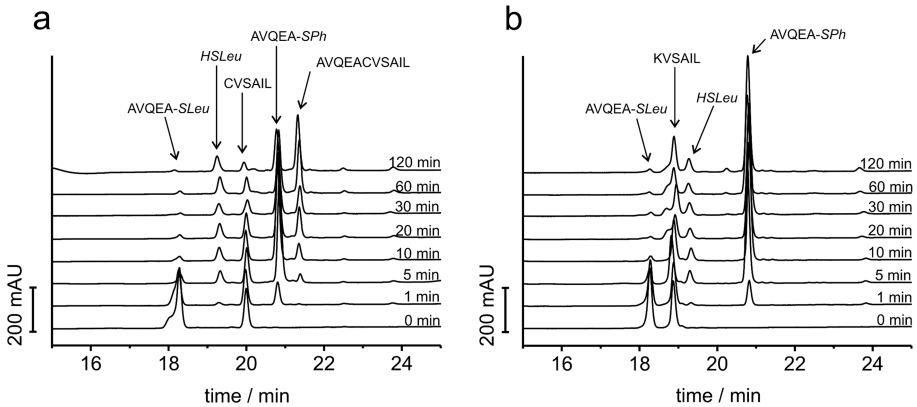

Figure 3. HPLC analysis of the NCL at different time points. Peaks were collected, analyzed with ESI-MS, and assigned. (a) Ligation with Cys-Val-Ser-Ala-Ile-Leu and (b) ligation with Lys-Val-Ser-Ala-Ile-Leu as C-terminal peptide (S*Leu* = $-SCH_2CO$*Leu*).

only Ala-Val-Gln-Glu-Ala-S*Ph* is generated indicating that primary and ε-amino groups are not reacting or are forming only in minor yields.

To determine whether such side reactions do occur, more flat gradients were used. Figure 4 summarizes the results. Data were processed as described in the Materials and Methods section. The fraction of educts and products was plotted against the reaction time at 0 and 2 h. It is evident that small amounts ($<$2%) of the thioester are reacting with primary amines in both experimental systems using thiophenol or MPAA. On the other hand, the classical NCL with an N-terminal Cys is kinetically favored so that side products are not formed (detection limit $<$0.1%). Under these conditions, a yield of about 50% of the final NCL product is obtained after 2 h reaction time. The conversion of Ala-Val-Gln-Glu-Ala-S*Leu* to the corresponding aryl thioester can be observed in all three experiments. It should be noted that dimerization product **4** has not been detected in the HPLC chromatograms. However, careful mass spectrometric analysis of the crude reactions mixture indicated minute amounts.

In an additional set of experiments, the guanidinium group of Arg was tested as possible reaction partner. With the use of both thiol reagents (thiophenol and MPAA), commercially available N-terminal Z-protected Arg was allowed to react with the N-terminal peptide (Ala-Val-Gln-Glu-Ala-S*Leu*). No reaction product could be detected indicating that also the guanidinium group is inert towards reaction with N-terminal thioester (data not shown). It should be noted that some sequences might contain a high number of positive charges. In this case, a careful increase of base should be considered. As a summary, NCL can be performed chemoselectively in organic solvents without the need for protection of side chain functional groups.

Epimerization Test

Another potential side reaction of NCL in organic solvents concerns epimerization during the ligation reaction. This unwanted side reaction might arise in the presence of a base (e.g. TEA) as a result of a lower pK_a of the α–C–H bond of the C-terminal thioester.

To analyze this question, thiophenol and MPAA were used respectively for the synthesis of a short tripeptide. The N-terminal educt was an Ala-thioester and the C-terminal educt the dipeptide Cys-Val. The two possible epimers (D-Ala-Cys-Val and L-Ala-Cys-Val); can be separated by HPLC on a C18 column with a standard water/acetonitrile buffer system (Figure 5, trace a). As can be seen in both examples using MPAA and thiophenol as activating agents, no D-epimer is detected (detection limit is $<$0.1%).

To estimate whether epimerization can occur at all, Ala-S*Leu* was preincubated with thiophenol or MPAA and base for 2 h (Table 2). Subsequently, the Cys-Val peptide was added and

Journal of PeptideScience

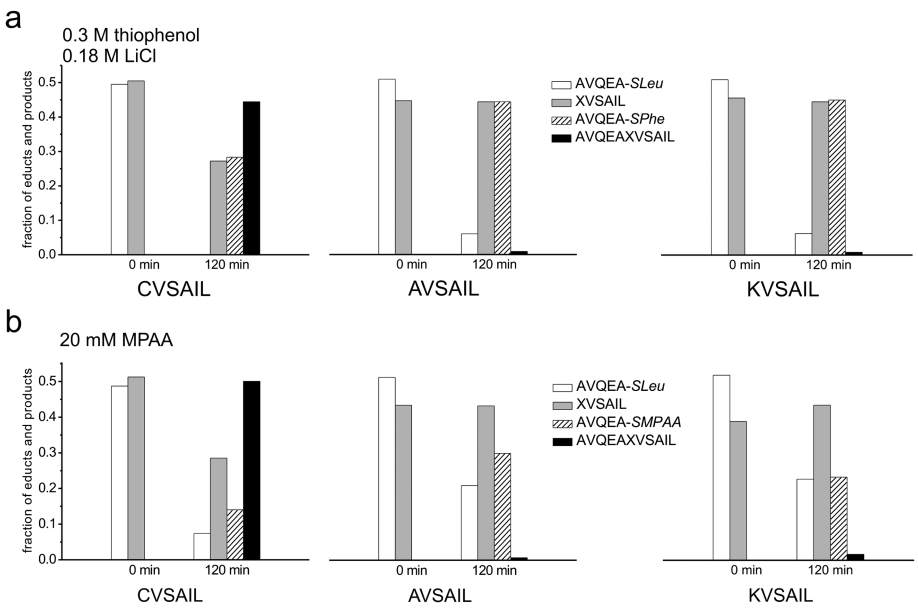

Figure 4. NCL in DMF. Plotted are the fractions of reaction compounds. Their sum is normalized to 1. (a) Reaction conditions: 0.3 M thiophenol, 0.18 M LiCl in DMF; (b) reaction conditions: 20 mM MPAA in DMF. (*SLeu* = $-SCH_2CO$*Leu*; X = Cys, Ala, or Lys). The yield (determined from three independent experiments) for the synthesis of Ala-Val-Gln-Glu-Ala-Cys-Val-Ser-Ala-Ile-Leu (4A left panel) was 46.1 ± 2.6%.

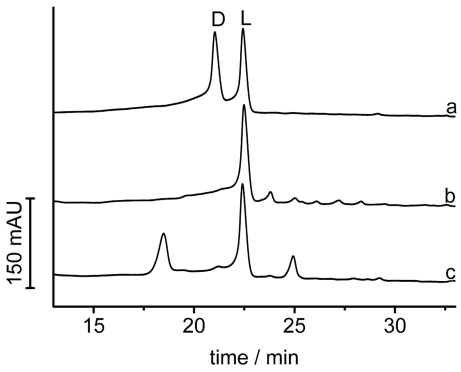

Figure 5. Racemization of NCL in DMF studied by the synthesis of the tripeptide Ala-Cys-Val: (a) D-Ala-Cys-Val and L-Ala-Val-Cys synthesized by SPPS; (b) NCL with MPAA (20 mM); (c) NCL with thiophenol (0.3 M thiophenol). The peaks at 17.5 and 25 min were analyzed by ESI-MS, but no mass could be identified. The all L-peptide is labeled with L and the D-Ala_1 epimer with D.

the reaction was stopped after 1 h by the addition of 0.1% TFA/water. After removal of organic solvent under reduced pressure the reaction mixture was analyzed by HPLC.

Table 2. Racemization studied by the synthesis of the tripeptide Ala-Cys-Val. TEA (20 mM) was used in all experiments

Reaction condition	L-Ala-Cys-Val (%)	D-Ala-Cys-Val (%)
20 mM MPAA	97	3
20 mM thiophenol	97	3
20 mM thiophenol/0.18 M LiCl	89	11
0.3 M thiophenol	64	36
0.3 M thiophenol/0.3 M benzyl mercaptan	56	44
0.3 M thiophenol/0.18 M LiCl	53	47

It is clearly evident that relatively low amounts of the D-epimer (D-Ala-Cys-Val) are formed, if the thiol additive does not exceed 20 mM. Higher concentrations (0.3 M thiophenol or 0.3 M thiophenol/benzyl mercaptan) lead to a considerable increase of D-epimer. This result can be explained by the high excess of the thiol additive, which leads to a fast and almost complete transesterification within 10 min (Figure 3) leaving enough time for racemization. Obviously, arylthioester is quite susceptible to racemization, but this reaction can be avoided by reducing the thiol concentration and mixing the reactants simultaneously.

To analyze whether these results hold true also for longer peptides, Ala-Val-Gln-Glu-Ala-*SLeu* and Cys-Val-Ser-Ala-Ile-Leu were used for NCL. The reference compounds Ala-Val-Gln-Glu-L-Ala-Cys-Val-Ser-Ala-Ile-Leu and Ala-Val-Gln-Glu-D-Ala-Cys-Val-Ser-Ala-Ile-Leu were synthesized by SPPS. These two epimers were separated on HPLC using a phosphate/acetonitrile buffer system (Figure 6, trace a). Formation of the epimers during ligation of Ala-Val-Gln-Glu-Ala-*SLeu* with Cys-Val-Ser-Ala-Ile-Leu was analyzed after the reaction was stopped at 2 h.

It is evident that only the all L-epimer is observed (detection limit is <0.1%) in agreement with the results obtained from the formation of the tripeptide Ala-Cys-Val. It should be noted that higher thiophenol concentrations do not lead to epimerization (Figure 6, trace c).

Concluding Remarks

In this work, we could demonstrate that NCL in DMF proceeds chemoselectively. Other functional amino groups present in the peptide chain (N-terminal amino group, Lys and Arg) are inert under these conditions. Importantly, NCL in organic solvents leads to the all L-epimer if both educts and thiol additives are mixed simultaneously. These results and the observation that the reaction proceeds quite fast and in high yields [12] open the way for the ligation of hydrophobic peptides and, ultimately, for the synthesis of membrane proteins.

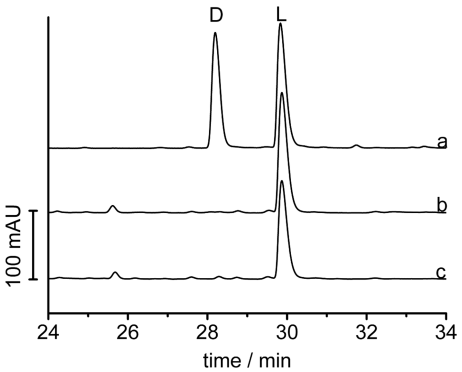

Figure 6. Racemization of NCL in DMF studied by the synthesis of Ala-Val-Gln-Glu-Ala-Cys-Val-Ser-Ala-Ile-Leu after a reaction time of 2 h. (a) Ala-Val-Gln-Glu-D-Ala-Cys-Val-Ser-Ala-Ile-Leu and Ala-Val-Gln-Glu-L-Ala-Cys-Val-Ser-Ala-Ile-Leu synthesized by SPPS. The all L-peptide is labeled with L and the D-Ala_5 epimer with D; (b) NCL with MPAA (20 mM MPAA, 20 mM TEA in DMF, 40 °C); (c) NCL with thiophenol (0.3 M thiophenol, 20 mM TEA, 0.18 M LiCl in DMF, 40 °C). Peptides were separated by HPLC on a C18 column using linear gradients from buffer A (5 mM disodium hydrogen phosphate, pH 8 in water) to B (5 mM disodium hydrogen phosphate, pH 8 in 60% (v/v) acetonitrile/40% (v/v) water).

Acknowledgements

We thank the 'Volkswagenstiftung' and the 'Studienstiftung des deutschen Volkes' for their financial support.

References

1 Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Synthesis of proteins by native chemical ligation. *Science* 1994; **266**: 776–779.
2 Kent SBH. Total chemical synthesis of proteins. *Chem. Soc. Rev.* 2009; **38**: 338–351.
3 Kumar KSA, Bavikar SN, Spasser L, Moyal T, Ohayon S, Brik, A. Total chemical synthesis of a 304 amino acid K48-linked tetraubiquitin protein. *Angew. Chem. Int. Ed.* 2011; **50**: 6137–6141.
4 Johnson ECB, Kent SBH. Towards the total chemical synthesis of integral membrane proteins: a general method for the synthesis of hydrophobic peptide-α-thioester building blocks. *Tetrahedron Lett.* 2007; **48**: 1795–1799.
5 Olschewski D, Becker CFW. Chemical synthesis and semisynthesis of membrane proteins. *Mol. Biosyst.* 2008; **4**: 733–740.
6 Kochendoerfer GG, Salom D, Lear JD, Wilk-Orescan R, Kent SBH, DeGrado WF. Total chemical synthesis of the integral membrane protein influenza A virus M2: role of its C-terminal domain in tetramer assembly. *Biochemistry* 1999; **38**: 11905–11913.
7 Valiyaveetil FI, MacKinnon R, Muir TW. Semisynthesis and folding of the potassium channel KcsA. *J. Am. Chem. Soc.* 2002; **124**: 9113–9120.
8 Sato T, Saito Y, Aimoto S. Synthesis of the C-terminal region of opioid receptor like 1 in an SDS micelle by the native chemical ligation: effect of thiol additive and SDS concentration on ligation efficiency. *J. Pept. Sci.* 2005; **11**: 410–416.
9 Clayton D, Shapovalov G, Maurer JA, Dougherty DA, Lester HA, Kochendoerfer GG. Total chemical synthesis and electrophysiological characterization of mechanosensitive channels from *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* 2004; **101**: 4764–4769.
10 Lahiri S, Brehs M, Olschewski D, Becker CF. Total chemical synthesis of an integral membrane enzyme: diacylglycerol kinase from *Escherichia coli*. *Angew. Chem. Int. Ed.* 2011; **50**: 3988–3992.
11 Musiol HJ, Dong S, Kaiser M, Bausinger R, Zumbusch A, Bertsch U, Moroder L. Toward semisynthetic lipoproteins by convergent strategies based on click and ligation chemistry. *Chembiochem* 2005; **6**: 625–628.
12 Dittmann M, Sauermann J, Seidel R, Zimmermann W, Engelhard M. Native chemical ligation of hydrophobic peptides in organic solvents. *J. Pept. Sci.* 2010; **16**: 558–562.
13 Klare JP, Chizhov I, Engelhard M. Microbial rhodopsins: scaffolds for ion pumps, channels, and sensors. In *Bioenergetics Results and Problems in Cell Differentiation (45) Bioenergetics: Energy Conservation and Conversion*, Schäfer G, Penefsky H (eds). Springer-Verlag: Berlin Heidelberg, 2008; 73–122.
14 Seebach D, Thaler A, Beck AK. Solubilization of peptides in non-polar organic solvents by the addition of inorganic salts: facts and implications. *Helv. Chim. Acta* 1989; **72**: 857–867.
15 Robey FA. Selective and facile cyclization of N-chloroacetylated peptides from the C4 domain of HIV Gp120 in LiCl/DMF solvent systems. *J. Pept. Res.* 2000; **56**: 115–20.
16 Schnölzer M, Alewood P, Jones A, Alewood D, Kent SBH. *In situ* neutralization in Boc-chemistry solid phase peptide synthesis. *Int. J. Pept. Protein Res.* 1992; **40**: 180–193.
17 Hackeng TM, Griffin JH, Dawson PE. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. USA* 1999; **96**: 10068–10073.

