Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 14 (2006) 5043–5048
Bioorganic & Medicinal Chemistry

Expression of N-terminal Cys-protein fragments using an intein refolding strategy

Christian P. R. Hackenberger,[†] Mark M. Chen and Barbara Imperiali[*]

Massachusetts Institute of Technology, Department of Chemistry and Department of Biology, 77 Massachusetts Ave. 18-590, Cambridge, MA 02139, USA

Received 31 January 2006; revised 1 March 2006; accepted 2 March 2006
Available online 22 March 2006

Abstract—The inclusion body expression and refolding of a pH-sensitive intein fusion protein (Ssp DnaB intein) delivered sufficient quantities of an N-terminal Cys-polypeptide for native chemical ligations. This strategy circumvents premature intein cleavage under expression conditions and allows the expression and purification of proteins with uncertain solubility properties. The expressed protein resembles the C-terminal portion of the amphiphilic immunity protein Im7, which can be ligated to synthetic thioesters to yield synthetic protein analogues for protein folding studies.

1. Introduction

Expressed protein ligation (EPL) has been established as a powerful tool for accessing quantities of pure proteins for biophysical and biochemical investigations.[1] The strategy employs a combination of chemical (solid-phase peptide synthesis, SPPS) and biochemical (expression) methods to provide protein fragments for connection via native chemical ligation (NCL).[2] NCL is based on a reversible thioester exchange process involving a protein or peptide thioester **1** and an N-terminal (or α-) Cys-protein or -peptide **2**,[3] which ultimately leads to irreversible amide bond formation (Scheme 1). EPL has found particular utility for the preparation of semisynthetic proteins, since it allows introduction of a variety of chemical modifications into the synthetic peptide fragment. Additionally, EPL affords access to membrane proteins that show low expression profiles, as shown by the semisynthesis of a potassium ion channel,[4] and it can even be performed in living cells.[5]

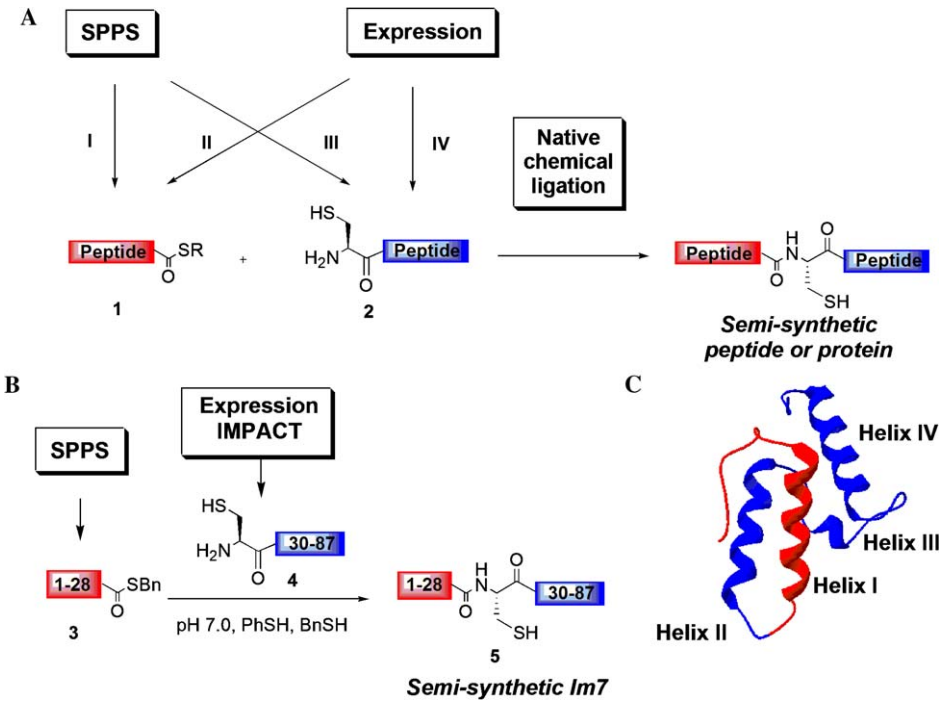
For access to semisynthetic proteins, both precursors for the NCL step have to be readily available. SPPS methods are well established and afford good quantities of synthetic peptide thioesters **1** or peptides with an N-terminal Cys **2** for the NCL step (routes I and III in Scheme 1A). For the expression and purification of analogous protein fragments for NCL, different routes for each type of protein component are available.

For the expression of protein thioesters **1** (route II in Scheme 1A) methods have been developed through the intein approach (IMPACT), which utilizes the self-cleaving property of a protein splicing domain or intein, in order to separate a desired polypeptide sequence from an affinity tag ('C-terminal fusion,' Scheme 2, left).[6] Protein splicing is then initiated once the fusion protein is loaded onto an affinity column by the addition of thiols to liberate the corresponding thioester in pure form. This protocol has been widely used to access soluble and insoluble protein domains, the latter as recently demonstrated by Muir and co-workers in the semisynthesis of a potassium ion channel, in which the C-terminal intein fusion protein was expressed as inclusion bodies and refolded before the intein-mediated rearrangement was initiated.[4]

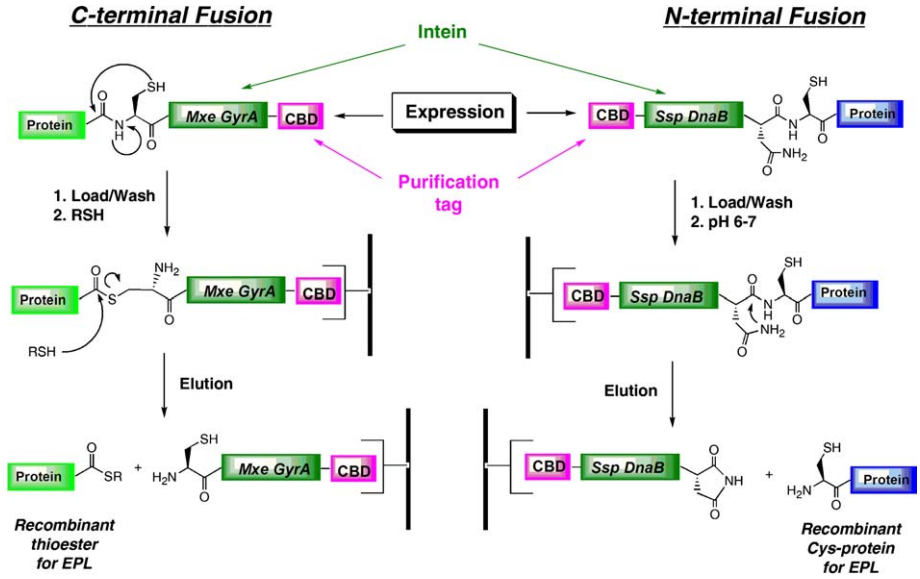
Soluble N-terminal Cys proteins **2** (route IV in Scheme 1A) are commonly expressed with a purification tag and a protease cleavage site for factor Xa[7] or TEV proteases[8] at the N-terminus, which allows the liberation of the protein.[1a] In order to access hydrophobic α-Cys-proteins that are insoluble or aggregate under protease

Keywords: Protein semisynthesis; Intein expression; N-terminal Cys-proteins; Native chemical ligation.
* Corresponding author. Tel.: +1 617 253 1838; fax: +1 617 452 2419; e-mail: imper@mit.edu
† Present address: Freie Universität Berlin, Department of Chemistry and Biochemistry, Takustr. 3, 14195 Berlin, Germany.

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.03.003



Scheme 1. (A) Principle of expressed protein ligation (EPL). (B) Semisynthetic route toward the immunity protein Im7. (C) Structure of the immunity protein Im7.



Scheme 2. Recombinant strategies toward N-terminal Cys-proteins and protein-thioesters using the intein approach (IMPACT). CBD, chitin-binding domain; *Mxe GyrA* and *Ssp DnaB*, intein domains.

cleavage conditions, cyanogen bromide (CNBr)-mediated cleavage between a Met-Cys motif in a protein has recently been reported.[9]

Another valuable approach for the expression and purification of recombinant α-Cys-proteins **2** would be intein cleavage, analogous to the thioester expression. In this route, referred to as 'N-terminal fusion,'[10] the intein is derived from the *Synechosystis sp dnaB* gene (*Ssp DnaB*) and undergoes pH or temperature dependent cleavage (Scheme 2, right). However, in contrast to the C-terminal fusion, this method has been rarely used due to the occurrence of premature intein cleavage under the expression conditions, which are often slightly acidic and thus lead to an immediate release of the purification tag.[1a]

Herein, we present an expression strategy that circumvents premature intein cleavage and provides a general approach for access to hydrophobic or amphiphilic α-Cys-proteins **2** for EPL. This paper defines the obstacles encountered in soluble intein expression and

delineates a protocol which implements inclusion body expression, isolation, and subsequent refolding.[11] This is applied toward the preparation of the amphiphilic C-terminal sequence (29–87), **4**, of the four-helical bundle protein Im7 (Scheme 1B).[12] This protein fragment can then be ligated to synthetic N-terminal peptide thioesters **3** to afford the semisynthetic Im7 variants for comparative protein folding studies (Scheme 1B).[13]

2. Results and discussion

Considering the structure of native Im7[14] we anticipated that the expressed fragment 29–87 of Im7, lacking helix I, would expose a hydrophobic area of the protein core (Scheme 1C), thus the expression of this protein fragment represented a good test case for the development of methodology to access hydrophobic or amphiphilic α-Cys-proteins. In initial studies, we focused on the N-terminal intein approach to access (Im7) Ala29Cys-87, **4**, despite the potential obstacle of premature cleavage.[1a] It should also be noted that it had been reported that temperature optimizations were useful for enabling the method,[15a,b] and that recent studies have provided expressed proteins after refolding of fusion proteins from inclusion bodies.[4,11b] Additionally, we hoped that the fusion protein CBD–intein–Ala29Cys-87, in which the CBD–intein part constitutes 75% of the protein mass, would lead to improved solubility properties and thereby enabling expression and purification.

2.1. Cloning and initial expression studies

The DNA fragment corresponding to the Ala29Cys-87 peptide of Im7 fused to a pH-labile intein and a chitin-binding domain (CBD) (compare Scheme 2, right) was inserted into the IMPACT-pTWIN1 vector and transformed into *Escherichia coli* (see Supplementary material).[16] A pilot expression for 2 h at 37 °C was performed. Peptide gel electrophoresis in combination with Western blot analysis against a CBD-antibody showed that the desired fusion protein CBD–intein–Ala29Cys-87 was only expressed in the cell lysate to a minor extent (Fig. 1, lane 2 in peptide gel and Western blot). Under the expression conditions, the major product was the 'cleaved' CBD–intein fragment (25 kDa), which had already undergone the intein cleavage thus liberating the Ala29Cys-Gly87 peptide **4**. We accounted for this problem by the fact that expression conditions commonly result in a pH decrease, as in our case to ~pH 6.8, which initiates the intein cleavage and fails to deliver the protein for the on-column purification step.[1a]

2.2. Refolding of the intein fusion protein

At this point extensive studies were performed to optimize the expression conditions for the 'uncleaved' protein, while suppressing the intein cleavage. Numerous parameters, such as lowering the expression temperature to 16 °C, reducing the expression time to 0.5–1 h, and increasing the pH of the expression medium to 7.5, were investigated; however, none of the conditions led to significant amounts of the desired fusion protein CBD–intein–Ala29Cys-87 in the cell lysate for the on-column purification step.

Reevaluation of the initial test cleavage (Fig. 1) finally revealed a solution. As visible in both gels, the resuspended cells contain the desired 'uncleaved' protein (upper band at 36 kDa) as the major product. This peptide is presumably accumulated in denatured form in inclusion bodies and thus lacks the ability to mediate the intein cleavage. We isolated the inclusion bodies under standard conditions (7 M guanidinium buffer, see Section 4) and subjected this material to refolding conditions to reestablish the functionality of the intein cleavage domain.[17] The refolding strategy employed stepwise dialysis against decreasing concentrations of urea at pH 8.5.[18] Refolding and protein purification were monitored by native peptide gel analysis (Fig. 2A). The dialysis product was added at pH 8.5 onto a chitin column (Fig. 2A, lane 2) and purified by iterative washing at pH 8.5 (lanes 3–6). On-column intein cleavage for 16 h was initiated by dropping the pH to 6.0, however, only very small amounts of peptide products were observed (lane 7).[19] Further intein cleavage for 24 h by elution with a 50 mM solution of DTT at pH 6.0 increased the yield significantly and the Ala29Cys-87 peptide **4** was observed as the major product (lane 8). Increasing the cleavage time for another 16 h only led to a small

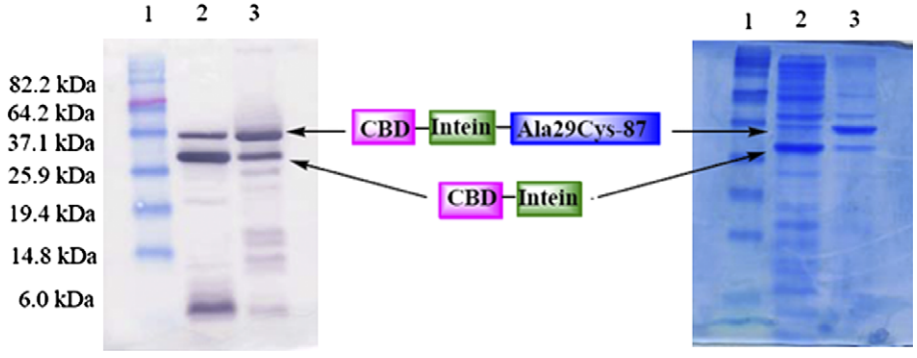


Figure 1. Pilot expression of fusion protein CBD–intein–Ala29Cys-87. Western blot (left) and stained gel (right): lane 1, molecular weight markers; lane 2, cell lysate (20 μL from 100 mL cell lysis fraction loaded); lane 3, resuspended cells (20 μL from 100 mL resuspended cell fraction loaded). For Western blot analysis conditions, see Section 4.

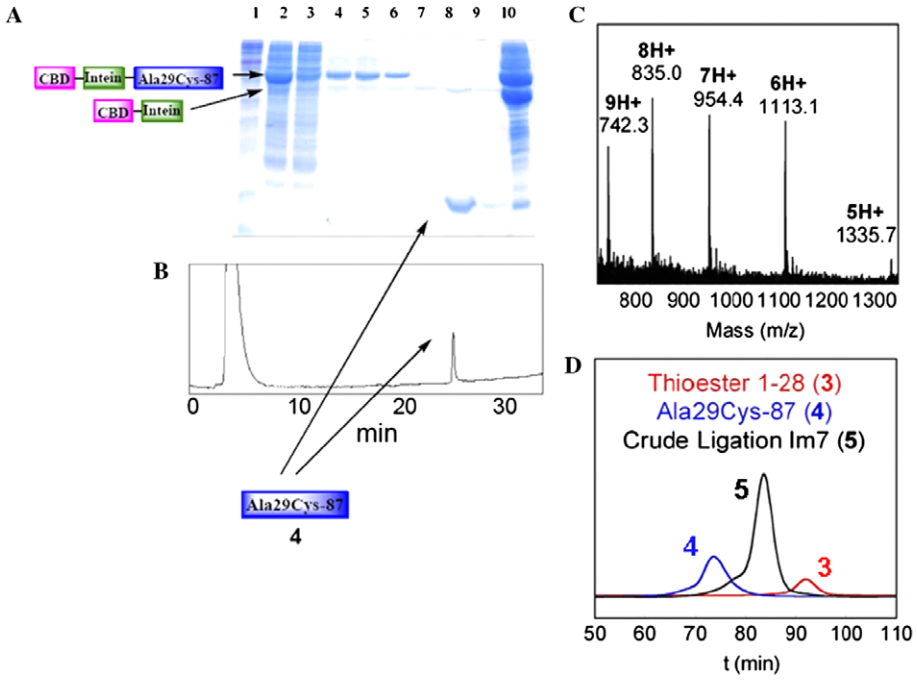


Figure 2. (A) Elution of CBD column and intein-mediated cleavage. Lane 1, molecular weight markers; lane 2, cell lysate after dialysis; lane 3, chitin column flowthrough; lane 4, wash (1) at pH 8.5; lane 5, wash (2) at pH 8.5; lane 6, wash (3) at pH 8.5; lane 7, elution at pH 6.0 (after 12 h); lane 8, elution at pH 6.0 with DTT (after 36 h); lane 9, elution at pH 6.0 with DTT (after 52 h); lane 10, boiled beads (after 52 h). (B) HPLC spectrum of crude Ala29Cys-87 (**4**) after intein-mediated cleavage (lanes 8 and 9 in (A)). Gradient for HPLC: 7–95% CH_3CN with 0.1% TFA over 35 min; Flow rate: 1 ml/min, C_{18}-column. (C) ESI-MS of Ala29Cys-87 (**4**). (D) SEC of crude NCL of **3**. See Section 4 for SEC conditions.

amount of additional peptide product (lane 9). A final analysis of the chitin beads revealed that about 50% of the fusion protein was cleaved (lane 10).

The fractions containing the desired protein fragment were collected, concentrated by lyophilization, and analyzed by HPLC (Fig. 2B) and ESI-MS (Fig. 2C). The crude material was found to be >90% pure and could be further purified by preparative HPLC separation.

Based on additional optimization studies, we made several important observations. First, we found that long dialysis times during each refolding step (each ~12 h) increased the intein cleavage significantly, probably due to a higher content of correctly refolded fusion proteins, which are capable of intein rearrangement. Second, the resuspended cells of a 1-L expression have to be dialyzed against a minimum volume of 1-L refolding buffer. Third, the on-column cleavage must be performed under reducing conditions, which we rationalize is due to the reduction of disulfides, which could otherwise lead to a decrease in the intein cleavage yield.

With the application of these conditions a typical 1-L expression yields about 3–4 mg of pure peptide, which represents an excellent expression yield since only one-quarter of the originally expressed protein is isolated.

2.3. NCL and aggregation observed during SEC separation

Having sufficient quantities of the expressed protein Ala29Cys-87 **4** in hand, we were able to access various semisynthetic Im7 analogues by ligation with synthetic peptide thioesters (Scheme 1B). During the NCL of expressed **4** with thioester **3** for the semisynthetic Im7 **5** containing a single Ala29Cys mutation, we noted that the expressed protein Ala29Cys-87 **4** (6.7 kDa) showed a tendency to aggregate (as evidenced by elution before the full-length Im7-protein **5** (10.1 kDa)) during size-exclusion chromatography (SEC, Fig. 2D). This behavior supports initial predictions concerning the properties of the expressed Ala29Cys-87 fragment.

3. Conclusion

In summary, we describe a protocol for the expression of N-terminal Cys-proteins based on a pH-sensitive intein fusion protein, which includes the refolding of the *Ssn DnaB* intein after inclusion body isolation and thus circumvents the premature intein cleavage. We propose that this methodology offers an alternative approach for the expression of proteins with uncertain solubility properties that are inaccessible by protease cleavage. In addition, this method does not face many of the limitations in the expressed protein sequence as encountered with proteases or CNBr. However, it should be noted that this method requires inclusion body expression and subsequent refolding of intein fusion proteins, which might limit the number of suitable intein fusion proteins that can be used. Further attempts to generalize the intein methodology and to investigate the influence of the intein domain on the formation of inclusion bodies as well as folding studies of the semisynthetic immunity proteins are currently underway.

4. Experimental

4.1. General

All reagents, amino acids, and solvents were purchased from commercial suppliers and used without further purification. Urea for the dialysis buffers was purchased from Aldrich (98%). All solvents were of reagent grade and used as received. Procedures for the synthesis of Im7 thioesters as well as conditions and work-up for the native chemical ligations were described previously.[13]

The following abbreviations were used:

NCL, native chemical ligation; EPL, expressed protein ligation; SEC, size-exclusion chromatography; LB-medium, Luria–Bertani medium; IPTG, isopropyl β-D-1-thiogalactopyranoside; CBD, chitin-binding domain.

The following buffers were used during the investigations: cell lysis buffer: 20 mM Tris–HCl, pH 8.0, and 0.5 M NaCl. Breaking buffer: cell lysis buffer, 7 M guanidine–HCl, and 10 mM DTT. Refolding buffer A: cell lysis buffer, 8 M urea, and 10 mM DTT. Refolding buffer B: cell lysis buffer, 6 M urea, and 1 mM DTT. Refolding buffer C: cell lysis buffer, 4 M urea, and 1 mM DTT. Refolding buffer D: cell lysis buffer, 2 M urea, and 1 mM reduced glutathione. Refolding buffer E: cell lysis buffer, 1 mM reduced glutathione. Buffer A: 20 mM Tris–HCl, pH 8.5, and 0.5 M NaCl. Buffer B: 20 mM Tris–HCl, pH 6.0, 0.5 M NaCl, and 50 mM DTT.

4.2. Cloning

The protocol for the insertion of the Im7-DNA fragment in the pTWIN vector and transformation into a host cell line was adapted from the standard procedure provided by New England Biolabs.[15] Detailed experimental procedures can be found in Supplementary material.

4.3. Expression

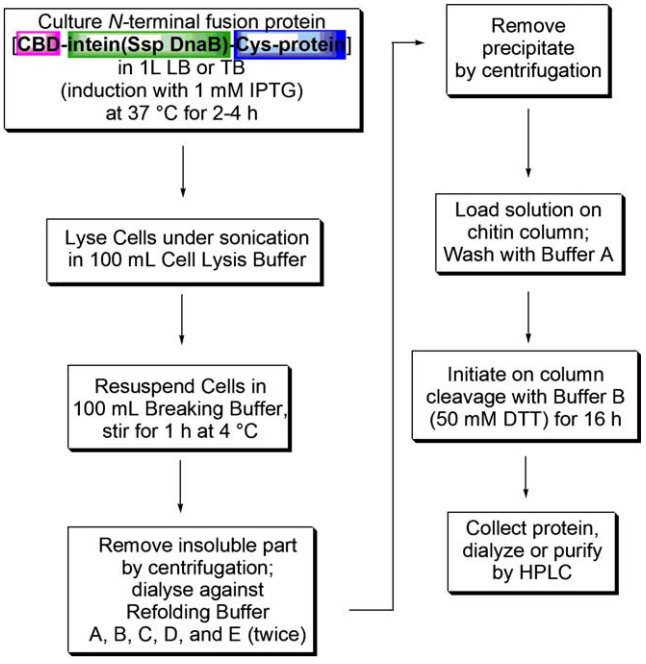
Expression studies of the CBD–intein–Ala29Cys-87 fusion protein were performed in LB-medium containing 1 μM carbenicilin. Cultures were grown until an OD value of 0.90 was reached before IPTG was added for a final concentration of 1 mM. After expression, the cells were spun down (12,000*g*), resuspended in 100 mL cell lysis buffer, lysed by sonication and again spun down. The remaining pellet was resuspended ('resuspended pellet') in 100 mL cell lysis buffer.

4.4. Optimization studies

Expression profiles were monitored by peptide gel and Western blot analysis. For Western blot analysis the membranes were treated with a mouse anti-CBD primary antibody followed by a goat anti-mouse antibody conjugated with alkaline phosphatase and then visualized with a chromogenic alkaline phosphatase substrate (Fig. 1). During the initial pilot expressions, various parameters were changed including temperature (16, 25, and 37 °C), time (20 min to 6 h) and the pH of the expression medium (6.0–8.0), however, none of these studies delivered the intact fusion protein.

4.5. Inclusion body expression and refolding

For the isolation of the intein fusion protein as inclusion bodies, expression was initiated with 1 mM IPTG, carried out at 37 °C for 3 h, and worked up as described before. The denatured fusion protein was extracted from the lysed cells by suspension with the breaking buffer for 1 h at 4 °C (100 mL for a 1-L expression pellet). The suspension was spun down (15,000*g*) and the supernatant containing the unfolded fusion protein was refolded by stepwise dialysis against decreasing concentrations of urea (Scheme 3) using refolding buffers A, B, C, D, and E (twice).[16] Each dialysis step was performed at 4 °C for at least 12 h. About 1 L dialysis buffer was used for the refolding of a 1-L expression culture. After refolding, the dialysis product was added to a chitin column and the column was washed with 15 column volumes of buffer A. Intein cleavage was initiated by replacing buffer A with buffer B and the column was kept at pH 6.0 at room temperature overnight. The intein-cleaved peptide product Ala29Cys-87 **4** was finally eluted from the column by washing with buffer B. The protein product could further be purified by HPLC using a gradient of 7–95% CH_3CN with 0.1% TFA over 35 min. The desired peptide **4** eluted at 26.9 min and was stored in lyophilized form at −20 °C under nitrogen atmosphere until used for NCL experiments. The new refolding protocol for the purification and isolation of N-terminal Cys-proteins using the intein approach is illustrated in Scheme 3.



Scheme 3. General refolding protocol for the purification and isolation of N-terminal Cys-proteins using the IMPACT intein system.

4.6. Characterization of semisynthetic protein 5 and precursors

4.6.1. Thioester 1-28 (3). The peptide thioester **3** was synthesized as described previously[13] and purified via HPLC at a gradient of 7–95% CH_3CN over 35 min. Thioester **3** eluted at 24.0 min. MS (ESI-MS, see Figure S1 in Supplementary material): 1126.9 ($[M/3]^+$, observed), 1126.9 ($[M/3]^+$, calculated); 845.1 ($[M/4]^+$, observed), 845.4 ($[M/4]^+$, calculated).

4.6.2. Ala29Cys-87 (4). MS (ESI-MS, see Fig. 2C): 1335.7 ($[M/5]^+$, observed), 1335.7 ($[M/5]^+$, calculated); 1113.1 ($[M/6]^+$, observed), 1113.2 ($[M/6]^+$, calculated); 954.4 ($[M/7]^+$, observed), 954.3 ($[M/7]^+$, calculated); 835.0 ($[M/8]^+$, observed), 835.2 ($[M/8]^+$, calculated); 742.3 ($[M/9]^+$, observed), 742.5 ($[M/9]^+$, calculated).

4.6.3. Semisynthetic Im7 (5). The semisynthetic Im7 protein **5** was purified via SEC, which eluted at 80.0 min (See Fig. 2C, for exact conditions, see Ref. 13). MS (ESI-MS, see Figure S2 in Supplementary material): 1655.6 ($[M/6]^+$, observed), 1655.5 ($[M/6]^+$, calculated); 1419.1 ($[M/7]^+$, observed), 1419.2 ($[M/7]^+$, calculated); 1241.9 ($[M/8]^+$, observed), 1241.9 ($[M/8]^+$, calculated); 1103.9 ($[M/9]^+$, observed), 1104.0 ($[M/9]^+$, calculated); 993.8 ($[M/10]^+$, observed), 993.7 ($[M/10]^+$, calculated); 903.5 ($[M/11]^+$, observed), 903.5 ($[M/11]^+$, calculated); 828.2 ($[M/12]^+$, observed), 828.3 ($[M/12]^+$, calculated).

Acknowledgments

This work was funded by the NIH (GM 39334). C.P.R.H. acknowledges a postdoctoral fellowship from the German Academic Exchange Service (DAAD) and the German Science Foundation (DFG). We thank Elizabeth Vogel, Dr. Jebrell K. Glover, Dr. Guofeng Zhang, and Mary K. O'Reilly for valuable discussions and advice, Prof. Sheena Radford for the gift of the Im7-plasmid as well as Dr. Maurice Southworth and Dr. Thomas Evans, Jr. from New England Biolabs for suggestions concerning the intein refolding strategy.

Supplementary data

Full experimental procedures for cloning experiments as well as MS spectra of thioester **3** and semisynthetic protein **5**. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2006.03.003.

References and notes

1. (a) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249–289; (b) Casi, G.; Hilvert, D. *Curr. Opin. Struct. Biol.* **2003**, *13*, 589–594; (c) Hofmann, R. M.; Muir, T. W. *Curr. Opin. Biotechnol.* **2002**, *13*, 297–303.
2. (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776–779; (b) Bang, D.; Chopra, N.; Kent, S. B. H. *J. Am. Chem. Soc.* **2004**, *126*, 1377–1383.
3. Kemp, D. S.; Buckler, D. R. *Tetrahedron Lett.* **1991**, *32*, 3013–3016.
4. Valiyaveetil, F. I.; MacKinnon, R.; Muir, T. W. *J. Am. Chem. Soc.* **2002**, *124*, 9113–9120.
5. Giriat, I.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 7180–7181.
6. Evans, T. C., Jr.; Benner, J.; Xu, M.-Q. *J. Biol. Chem.* **1999**, *274*, 3923–3926.
7. Erlanson, D. A.; Chytil, M.; Verdine, G. L. *Chem. Biol.* **1996**, *3*, 981–991.
8. Tolbert, T. J.; Wong, C.-H. *Angew. Chem., Int. Ed.* **2002**, *41*, 2171–2174.
9. Macmillan, D.; Arham, L. *J. Am. Chem. Soc.* **2004**, *126*, 9530–9531.
10. Mathys, S.; Evans, T. C., Jr.; Chute, I. C.; Wu, H.; Chong, S.; Benner, J.; Liu, X.-Q.; Xu, M.-Q. *Gene* **1999**, *231*, 1–13.
11. For previous reports on the use of inclusion body expression and refolding, see: (a) Rudolph, R.; Lilie, H. *FASEB J.* **1996**, *10*, 49–56; (b) Guo, C.; Li, Z.; Shi, Y.; Xu, M.; Wise, J. G.; Trommer, W. E.; Yuan, J. *Protein Expr. Purif.* **2004**, *37*, 361–367.
12. (a) Capaldi, A. P.; Ramachandra Shastry, M. C.; Kleanthous, C.; Roder, H.; Radford, S. E. *Nat. Struct. Biol.* **2001**, *8*, 68–72; (b) Gorski, S. A.; Le Duff, C. S.; Capaldi, A. P.; Kalverda, A. P.; Beddard, G. S.; Moore, G. R.; Radford, S. E. *J. Mol. Biol.* **2004**, *337*, 183–193; (c) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. *Nat. Struct. Biol.* **2002**, *9*, 209–216.
13. Hackenberger, C. P. R.; Friel, C. T.; Radford, S. E.; Imperiali, B. *J. Am. Chem. Soc.* **2005**, *127*, 12882–12888.
14. Dennis, C. A.; Videler, H.; Paupitt, R. A.; Wallis, R.; James, R.; Moore, G. R.; Kleanthous, C. *Biochem. J.* **1998**, *333*, 183–191.
15. (a) Pezza, J. A.; Allen, K. N.; Tolan, D. R. *Chem. Commun.* **2004**, 2412–2413; (b) Humphries, H. E.; Christodoulides, M.; Heckels, J. E. *Protein Expr. Purif.* **2002**, *26*, 243–248.
16. IMPACT-TWIN Instruction Manual, Version 1.2, New England Biolabs Inc.
17. For similar refolding protocols, see Ref. 11b and: (a) Lyles, M. M.; Gilbert, H. F. *Biochemistry* **1991**, *30*, 613–619; (b) Evans, T. C., Jr.; Benner, J.; Xu, M.-Q. *Protein Sci.* **1998**, *7*, 2256–2264; (c) Marston, F. A. O.; Hartley, D. L. *Methods Enzymol.* **1990**, *182*, 264–276.
18. During our experiments we found that stepwise refolding was necessary to re-establish intein functionality. Neither loading the fusion protein under mild denaturation conditions (4 M urea, compare Ref. 15b) followed by removal of urea nor dialysis of the denatured fusion protein against a buffer containing no urea (compare Ref. 4) delivered material capable of intein cleavage.
19. In a separate experiment, a non-reducing peptide gel was performed after on-column cleavage at pH 6.0. This gel showed two peptide bands, which correlate with the desired Ala29Cys-87 peptide and the oxidized disulfide dimer (data not shown).