Biochemical and Biophysical Research Communications **254,** 147–151 (1999)
Article ID bbrc.1998.9913, available online at http://www.idealibrary.com on IDEAL®

Use of Benzyl Mercaptan for Direct Preparation of Long Polypeptide Benzylthio Esters as Substrates of Subtiligase

Ervin Welker and Harold A. Scheraga[1]
Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

Received November 25, 1998

Subtiligase, a double mutant of subtilisin, has been shown to be capable of joining together two unprotected peptide fragments, namely an activated peptide ester and a second peptide with a free N-terminal amino group. Inside cells, inteins are know to join peptide chains (exteins) by self-extrusion. The SC VMA1 intein was modified to undergo only *in vitro* N-terminal cleavage in the presence of small nucleophilic compounds, releasing the N-terminal extein. With a proper choice of the nucleophilic compounds it is shown that it is possible to generate long polypeptides, by molecular biology expression, with such an attached reactive ester which is an excellent substrate of the enzyme, subtiligase. This approach can successfully extend the current limit of the subtiligase-catalyzed fragment condensation method as well as provide another application of the recently discovered intein chemistry.

The incorporation of unnatural amino acids by chemical synthesis has been a useful tool for probing structure-function relationships in proteins (1). Despite considerable progress in solid phase peptide synthesis (SPPS), the chemical synthesis of peptides with 60 or more amino acids remains a significant challenge. Recently developed selective enzymatic or chemical ligation methods have permitted the total synthesis of proteins of over 100 amino acids.

Two peptide ligation techniques, which lead to formation of a peptide bond, have been shown to be particularly useful (2, 3). In one of them, Native Chemical Ligation (NCL), an unprotected synthetic peptide A, bearing a C-terminal α-thio ester, is reacted in a chemoselective manner with an unprotected peptide B containing an N-terminal cysteine residue(2). Thiol exchange yields an initial thio ester-linked intermediate which rearranges spontaneously to give a peptide bond at the ligation site joining the two peptide segments, regenerating the cysteine side-chain thiol in the process (Fig. 1). Syntheses of a number of proteins of moderate size by a rapid single-step ligation have been reported (2). Therefore, NCL is restricted to joining peptide segments at an X-Cys bond. In principle, X can be any amino acid but, for optimal ligation, β-branched amino acids (such as Val) should be avoided (4). Trifunctional amino acids (such as Asn), whose side chains participate in assisted hydrolysis, were reported to reduce the yield because of hydrolysis of the thio ester (on peptide A) (5). Actually, in most cases reported, X was Gly or Ala. In four studies, the C-terminal amino acid of peptide A was even mutated to Gly or Ala to facilitate the coupling reaction (2, 4, 6). Other work is under way to expand the range of sequences suitable for use as chemical ligation sites (7).

The enzymatic ligation technique, the Subtiligase Catalyzed Fragment Condensation (SCFC) method, which can accommodate different amino acids at the junction position, has proven to be a useful alternative to NCL(3). SCFC is especially useful with sequences that do not have a cysteine at the desired position of the protein, where introduction of an extra cysteine could disturb the protein conformation considerably. This approach exploits the unique properties of the enzyme subtiligase, a double mutant of subtilisin, which catalyzes the condensation reaction selectively between an activated peptide ester P and a second peptide Q with a free N-terminal amino group. In the first step of the catalysis, an active ester peptide fragment P acylates the cysteine of the active center of subtiligase. Next, the N-terminal amino group of the C-terminal peptide fragment Q attacks the peptide–enzyme complex, yielding the ligated peptide with a

[1] To whom correspondence should be addressed. Fax: 607-254-4700. E-mail: has5@cornell.edu.

0006-291X/99 $30.00

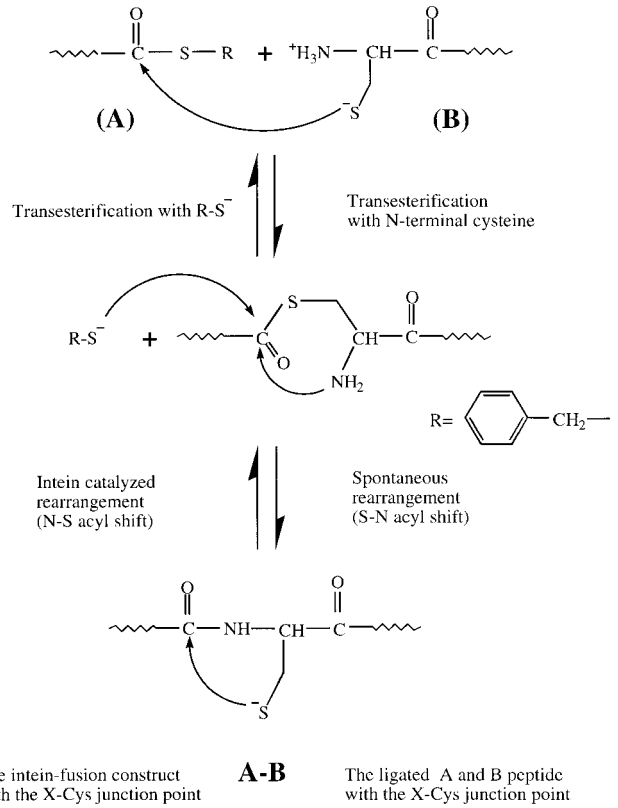

FIG. 1. Reactions involved in formation or cleavage of an X-Cys peptide bond. Because of the considerable similarity in the mechanisms, we illustrate the two reactions (NCL and chemical cleavage of the N-terminal extein) in the same figure. Down direction and right-hand notes: Mechanism of native chemical ligation. (A) N-terminal peptide thio ester. (B) C-terminal peptide containing N-terminal cysteine. (A-B) Ligated peptide formed by NCL. Up direction and left-hand notes: Simplified picture of the nucleophilic compound induced N-terminal cleavage of the mutated intein-fusion protein. In this case, A-B is the intein-fusion protein. (A) The liberated N terminal extein thio ester. (B) The intein.

peptide bond (Fig. 2). SCFC has been applied to total protein syntheses, semisyntheses and peptide cyclization (3).

Both methods allow the use of unprotected peptide segments, which avoids solubility problems and, therefore, increases the potential size of the synthesized protein significantly. In practice, however, the size is limited rather severely by the need to synthesize the peptide-esters (A or P) chemically. If molecular biology expression techniques could be used, it would be possible to overcome the "size barrier" of 200 amino acids.

The difficulties of preparing thio esters (or other activated ester derivatives) of a molecular-biologically expressed protein arise from the inability to distinguish the C-terminal carboxyl group from the side-chain carboxyls, and to modify it selectively in an unprotected peptide fragment. A generally applicable way to distinguish the C-terminal carboxyl group from the side-chain carboxyl groups has been found in intein chemistry (8). A modified protein splicing element [with an attached chitin-binding domain (CBD)] was used to establish a protein purification system which requires chemical cleavage of the intein fusion construct at the C-terminus of the protein to be expressed, without introducing extra amino acid(s), to serve as a recognition site for cleavage to liberate the desired peptide. This approach exploits a mutant variant of the SC VMA1 intein. Inteins are involved in the protein splicing process during which an internal sequence (the intein, which plays a role analogous to that of an intron in RNA splicing) catalyzes its self-extrusion from the protein by joining the lateral peptide chains (exteins) together. The mutant intein is not able to complete the splicing process but it can still catalyze the first step, the N-S acyl shift which produces a thio ester bridge between the C-terminus of the first extein and the N-terminal cysteine of the intein sequence (Fig. 1). In the presence of small nucleophilic molecules (e.g., hydrazine, reduced DTT, β-mercaptoethanol, cysteine), the thio ester bond is cleaved, releasing the N-terminal extein which is the peptide of interest (9). The CBD facilitates the easy purification of the expressed fusion protein using chitin beads (9).

The goal of our research is to establish a procedure, by improving the two methods above, to be capable of incorporating unnatural residues at any position in a

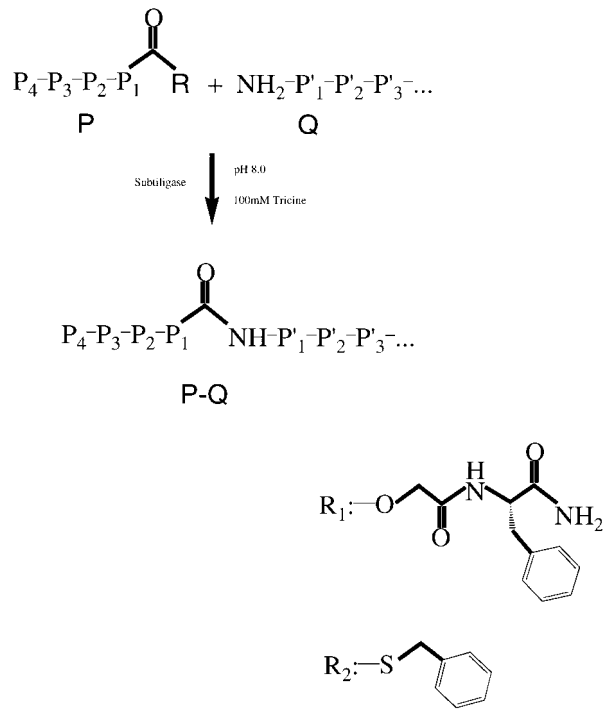

FIG. 2. Subtiligase catalyzed fragment condensation of peptides P and Q; R═R_1 for glycolate phenylalanyl ester, and R_2 for benzylthio ester.

protein of any size, even at multiple sites. This goal includes the development of a strategy for carrying out multiple ligation steps, which allows molecular-biologically expressed polypeptides to be used as the C- or N-terminal or intermediate fragments in these condensation reactions. As a part of this research, we examined the feasibility of using intein chemistry to produce the reactive ester derivative on peptide P for SCFC.

MATERIALS AND METHODS

Benzyl mercaptan was purchased from Aldrich, IPTG from Sigma, and chitin beads, DTT and plasmid MYB129 from New England Biolabs. Chong *et al.* (9) constructed the plasmid MYB129 to contain the coding sequences of the maltose binding protein-intein-chitin binding domain (CBD) fusion protein.

Escherichia coli strain BL21(DE3) harboring pMYB129 was used for fusion protein expression. One liter LB media supplemented with ampicillin (100 μg/ml) was inoculated with a freshly grown colony. The culture was incubated at 37°C until A_{600} = 0.6, and was then induced with isopropyl-β-D-thiogalactoside for 16 h at 22°C. The cells were harvested and disrupted by sonication in 5 ml Hepes buffer (20 mM Hepes, pH 7.6, 0.5M NaCl) on ice. After centrifugation, the crude supernatant was rotated slowly in 15-ml reaction tubes at 4°C with chitin beads which had previously been washed three times with 35 ml Hepes buffer (0.5 M NaCl). After 30 min incubation, the beads were centrifuged and washed three times with Hepes buffer (containing 0.5 M NaCl) and once with Hepes buffer (containing 0.05 M NaCl).

To induce the cleavage reaction, DTT (30 mM) or benzyl mercaptan [6% (v/v)], respectively, was added. The tubes were rotated slowly at room temperature for 1–3 h. After centrifugation, the supernatant and the first wash solution were diafiltered through a Microcon 30 filter (Amicon inc.) at 4°C. The protein in solution was stored frozen at −20°C and, after thawing, was analyzed by ESI-FT mass spectroscopy (10). Every step was monitored by SDS–PAGE. Alternatively, the protein was desalted by reverse-phase HPLC (Buffer A: water/0.2% acetic acid; buffer B: 50% acetonitrile/0.2% acetic acid; gradient 25–100%), the sample was concentrated with a Microcon 30 filter, and was analyzed by ESI-FT mass spectroscopy (10).

RESULTS AND DISCUSSION

The rate and success of the subtiligase ligation reaction depend on the nature of the substrate P used to acylate the enzyme. The reactive ester that is usually used is a glycolate phenylalanyl amide, which is sufficiently reactive, but not too reactive; therefore, the background hydrolysis of the peptide P ester does not become rate limiting(9). Although aromatic aryl ester and thioaryl ester substrates are more reactive than corresponding alkyl esters, aryl esters are more difficult to synthesize and are inherently less stable. Benzylthio esters, which are still more difficult to synthesize, satisfy the binding requirement (3) which is essential for effective acylation of the enzyme. The rate of ligation is about 20-fold higher than that for the glycolate esters (11). Therefore, we focussed on the preparation of a benzylthio ester with the intein-based expression system, with proper characterization of the product.

The 392-residue maltose binding protein was chosen to illustrate the method because the procedures for expression and purification of the maltose binding protein-intein-CBD are well established with the MYB129 vector (9). The expression and purification were carried out as in the Materials and Methods section, and the material was analyzed at every step by SDS–PAGE (data not shown). The fusion construct was expressed in a soluble form, with about 20 mg/l culture, in good agreement with data in the literature(9). The fusion protein bound selectively to the chitin beads, as shown by SDS–PAGE analysis (data not shown). The trans-esterification cleavage reaction was carried out in 50 mM NaCl, 20 mM Hepes, pH 8, 5–6% (v/v) benzyl mercaptan, at room temperature with slow rotation of the reaction tube for 1–3 h. The tubes were centrifuged, and the supernatant was found to contain only maltose binding protein which was purified by diafiltering with a Microcon 30 filter or, alternatively, by reversed phase HPLC. After the reversed phase separation, the sample was concentrated on a Microcon 30 filter in preparation for ESI-FT mass spectroscopy (carried out on two different purified samples which demonstrate that the maltose binding protein was cleaved to form the benzylthio ester with no detectable hydrolysis during the purification process). The resulting mass of 43,042 was in good agreement with the theoretical value (43,046, calculated with average isotope mass). The product was stored at −20°C and found to be stable, between pH 5 and 8, for several weeks, even during storage for several days at 4°C. A typical yield was about 10 mg of protein (terminated with a benzylthio ester) from 1 liter of culture.

This work demonstrates the high potential of the intein-based expression system for producing long polypeptide benzylthio esters in tens of mg quantities. Under our conditions, during 1–3 h at room temperature, the cleavage by benzyl mercaptan yields the thio ester without any decomposition. This thio-ester permits the synthesis, especially the incorporation of unnatural structures into the C-terminal segments, of proteins, much larger than previously achieved using SPPS-based enzymatic (SCFC) or chemical (NCL) methods.

For NCL reactions, the preparation of a benzylthio ester may not provide many advantages over the reported intein-based procedures (9). With NCL, to avoid difficulties in preparing and handling such an active ester, the phenyl thio ester is usually prepared *in situ* by trans-esterification of the original less active thio ester with 1.5% (v/v) thiophenol solution during the ligation reaction (2). Therefore, β-mercaptoethanol, which was also shown to initiate the cleavage reaction, works as well as benzyl mercaptan in NCL. Addition-

ally, due to steric hindrance around the active center of the intein, only small nucleophilic compounds were believed to be capable of performing the cleavage of the intein fusion construct (9, 12). For some amino acids preceding the cysteine at the junction of the intein fusion construct (X-Cys), even DTT, the most effective nucleophilic compound, was reported not to be able to induce cleavage (12). More recently, the impressive work of Muir and coworkers appeared (13). They introduced an extra Ala-Gly-Cys sequence as a junction point to facilitate thiophenol cleavage as well as the following ligation step. For sequences that are amenable to thiophenol cleavage, given the condition of the cleavage reaction of the intein-fusion construct and the NCL reaction, both steps can be completed simultaneously in the same reaction tube without the previous isolation of a reactive (even benzyl-) thio ester, as shown by Severinov and Muir (14).

For SCFC, however, the benzyl mercaptan cleavage provides more advantages over the reported intein based procedures (9, 13, 15). Subtiligase requires mild conditions for maintaining full activity, so that the prior isolation of the thio-ester for the ligation reaction is necessary (3). Isolation and handling of a benzylthio ester require less care than the very reactive phenylthio ester which is much more susceptible to hydrolysis (even during the SCFC reaction). In addition, the thiolate anion on its flexible intervening methylene group seems more likely to reach the carboxylate carbon at the cleavage site, which has been reported to be stericly hindered, even in cases of sequences which, due to greater steric hindrance, may not be amenable to thiophenol cleavage (7, 12). Therefore, the use of a benzylthio ester for SCFC is much more advantageous than a phenylthio ester. Theoretically, a long benzylthio ester could be prepared from the isolated product of the N-terminal cleavage reaction with smaller nucleophilic compounds [such as mercapto-acetic acid or β-mercaptoethanol (9), or sulfide which was reported recently (15)] in a second trans-esterification or (using sulfide) condensation step. In these cases, the second synthetic and additional purification steps may lower the yield considerably. Therefore, the use of benzyl mercaptan in the intein cleavage reaction that yields the benzylthioester in a single step (one hour incubation at room temperature, followed by a quick centrifugation), seems to be more advantageous.

We also consider the sequence requirements of the SCFC and the intein-based peptide preparation methods (3, 12). The amino acid X, preceding the cysteine at the junction position (X-Cys) of the SC VMA1 intein fusion construct, has a large impact on the efficiency of the cleavage reaction (8, 12). After the cleavage reaction, X becomes the C-terminal amino acid of the liberated peptide. Subtiligase is known to interact with the four C-terminal amino acids of peptide P: P_4-P_3-P_2-P_1. (3). Table 1 shows the preferred choices of the amino acids for both X and P_1. The number of amino acids that meet both criteria (with $X + P_1$ equal to +1 or +2) indicates the applicability of this intein-based method to prepare peptides with such an attached reactive ester derivative which is an excellent substrate for the enzyme, subtiligase.

TABLE 1

Amino Acid Preferences for Positions X and P_1 for Cleavage and Ligase Reaction Efficiency

	X[a]	P_1[b]	$X + P_1$[c]
Gly	1	−1	0
Ala	1	1	2
Val	0	−1	−1
Leu	1	1	2
Ile	0	−1	−1
Ser	1	0	1
Thr	0	−1	−1
Cys	−1	0	−1
Asp	−1	0	−1
Glu	0	0	0
Asn	−1	0	−1
Gln	1	0	1
Lys	1	1	2
Arg	−1	1	0
His	0	0	0
Phe	1	1	2
Tyr	1	1	2
Trp	1	0	1
Pro	−1	−1	−2
Met	1	1	2

[a] X, intein preference. Data from Ref. 12. The fraction of the protein that is not cleaved *in vivo* was multiplied by the fraction of the protein that was cleaved *in vitro* at 16°C. 1, above 0.8; 0, between 0.5 and 0.8; −1, under 0.5.

[b] P_1, subtiligase preference. Data from Ref. 1. −1, avoid; 0, intermediate; 1, best.

[c] The positive numbers (+1, +2) indicate the amino acids that meet both criteria.

ACKNOWLEDGMENTS

We thank M. J. Volles, W. J. Wedemayer, R. Zubarov, and F. W. McLafferty for help during this investigation. This work was supported by Grant GM-24893 from the National Institute of General Medical Sciences of the National Institutes of Health.

REFERENCES

1. Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J., and Wells, J. A. (1994) *Science* **266,** 243–247.
2. Muir, T. W., Dawson, P. E., and Kent, S. B. H. (1997) *Methods Enzymol.* **289,** 265–297.
3. Braisted, A. C., Judice, J. K., and Wells, J. A.(1997) *Methods Enzymol.* **289,** 298–313.
4. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. H. (1994) *Science* **266,** 776–779.

5. Tam, J. P., Lu, Y., Liu, C., and Shao, J. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 12485–12489.
6. Ueda, H., Siani, M. A., Gong, W., Thompson, D. A., Brown, G. G., and Wang, J. M. (1997) *J. Biol. Chem*. **272,** 24966–24970.
7. Canne, L. E., Ferre-D'Amare, A. R., Burley, S. K., and Kent, S. B. H. J. (1995) *J. Am. Chem. Soc.* **117,** 2998–3007.
8. Chong, S., Yang, S., Paulus, H., Benner, J., Perler, F. B., and Xu, M.-Q. (1996) *J. Biol. Chem*. **271,** 22159–22168.
9. 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., and Xu, M.-Q. (1997) *Gene* **192,** 271–281.
10. McLafferty, F. W. (1994) *Acc. Chem. Res.* **27,** 379–386.
11. Abrahmsen, L., Tom, J., Burnier, J., Butcher, K. A., Kossiakoff, A., and Wells, J. A. (1991) *Biochemistry* **30,** 4151–4159.
12. Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M-Q. (1998) *J. Biol. Chem*. **273,** 10567–10577.
13. Muir, T., W., Sondhi, D., and Cole,P., A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6705–6710.
14. Severinov, K., and Muir, T. W. (1997) *J. Biol. Chem*. **273,** 16205–16209.
15. Kinsland, C., Taylor, S. V., Kelleher, N. L., McLafferty, F. W., and Begley, T. P. (1998) *Protein Sci.* **7,** 1839–1842.