Expressed Protein Ligation of 5-Enolpyruvylshikimate-3-phosphate (EPSP) Synthase: An Application to a Protein Expressed as an Inclusion Body

Hak Jun Kim,[*] Hee Jae Shin,[†] Hyun Woo Kim,[‡] Sung-Ho Kang, and Young Tae Kim[§]

Department of Polar Applied Science, Korea Polar Research Institute, Incheon 406-840, Korea. [*]E-mail: hjkim@kopri.re.kr
[†]Marine Natural Products Laboratory, Korea Ocean Research & Development Institute, P.O. Box 29, Ansan 426-744, Korea
[‡]Department of Marine Biology and [§]Department of Microbiology, Pukyong National University, Busan 608-737, Korea
Received September 8, 2007

Expressed protein ligation (EPL) technique, joining recombinantly expressed proteins to polypeptides, has been widely adopted for addressing various biological questions and for drug discovery. However, joining two recombinant proteins together is sometimes difficult when proteins are expressed insoluble and unrefoldable, because ligation-active proteins *via* intein-fusion are obtainable when they are folded correctly. We overcame this limitation coexpressing target protein with additional methionine aminopeptidase (MAP) which enhances removal of the initiation methionine of recombinantly expressed protein. Our approach demonstrated that two domains of 46 kDa 5-Enolpyruvylshikimate-3-phosphate (EPSP) synthase, a target of herbicide glyphosate, were successfully joined by native chemical ligation, although its C-terminal domain was expressed as an inclusion body. The intein-fused N-terminal fragment of EPSP synthase (EPSPSN, residues 1-237) was expressed and the ligation-active thioester tagged N-terminal fragment (EPSPSN-thioester) was purified using a chitin affinity chromatography and mercapto-ethanesulphonate (MESNA) as intein thiolysis reagent. Its C-terminal fragment (EPSPSC, residues Met^{237}-238^{CYS}-427), expressed as an inclusion body, was prepared from an additional MAP-expressing strain. Protein ligation was initiated by mixing ~1 mM of EPSPSN-thioester with ~2 mM of $EPSPSC^{CYS}$ (residues 238^{CYS}-427). Also we found that addition of 2% thiophenol increased the ligation efficiency *via* thiol exchange. The ligation efficiency was ~85%. The ligated full-length EPSP synthase was dissolved in 6 M GdHCl and refolded. Circular dichroism (CD) and enzyme activity assay of the purified protein showed that the ligated enzyme has distinct secondary structure and ~115% specific activity compared to those of wild-type EPSP synthase. This work demonstrates rare example of EPL between two recombinantly expressed proteins and also provides hands-on protein engineering protocol for large proteins.

Key Words : Expressed protein ligation, Intein, 5-Enolpyruvylshikimate-3-phosphate synthase, Methionine aminopeptidase

Introduction

Expressed protein ligation (EPL) is a powerful method for ligating recombinantly expressed polypeptides in a chemo- and regioselective manner (see reviews[1-3]). EPL takes advantage of native chemical ligation and an intein which is destined to be removed out of internal protein domain as the result of protein splicing.[3,4] This protein engineering technique has made it possible to incorporate unnatural amino acids, various noninvasive biophysical probes and stable isotopes within proteins regiospecifically, and also to produce toxic proteins *via* chemical ligation.[5] This technique combined with spectroscopy could be very useful for the study of protein-protein interactions, signal transduction and drug screening.[6]

So far, a great number of applications of this technique correspond to the ligation of recombinantly expressed protein with relatively short polypeptides chemically synthesized on a solid phase, since physical probes can be rather easily incorporated during the chemical synthesis.[3] Very recently, using protein *trans*-splicing (PTS), novel protein ligation system was introduced and expanded the size of proteins or fragments to be ligated.[7] This approach uses naturally or artificially split inteins which associate as functional form and eventually whose flanking polypeptides will be linked by a polypeptide bond. The drawback of the intein approach is that intein-fusion protein should be soluble or refoldable if it is expressed as an inclusion body. Otherwise, obtaining ligation-active proteins or domains is very limited (Figure 1B). We encounter this solubility issue especially when we try to ligate recombinantly expressed large domains or proteins together. To overcome this limitation, we employed methionine aminopeptidase (MAP) which enhances removal of the initiation methionine, leaving the penultimate residue, in this case cysteine, exposed for the ligation reaction.

EPSP synthase, catalyzing the unusual transfer of an enolpyruvyl moiety from phosphoenolpyruvate (PEP) to the 5-OH of shikimate 3-phosphate (S3P), is located in the shikimate pathway, a biosynthetic route to aromatic amino acids in bacteria, algae, fungi and plants, as well as apicomplexan parasites.[8-10] The extensive kinetic, biochemical and structural studies on this protein have been actively driven by the fact that it is the primary site of action of the herbicide glyphosate, the active ingredient of Roundup™.[11,12] EPSP synthase during catalysis can be a target for new drug

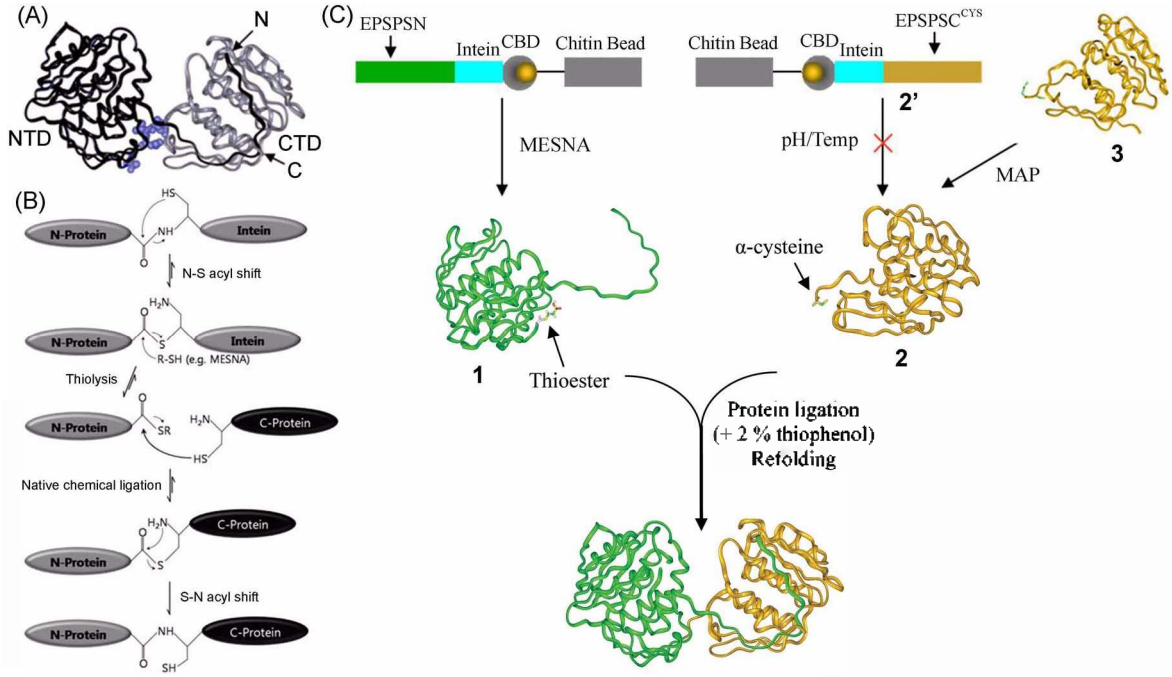

Figure 1. (A) Ribbon diagram of EPSP synthase. NTD and CTD are the N- and C-terminal domains respectively. The mutated residues (T236, L238, V239, and E240 from left to right) are displayed as CPK and in blue. (B) Chemistry of expressed protein ligation (EPL). For detailed description, see reviews.[1-3] N- and C-protein; N- and C-terminal fragment of protein of interest. (C) Schematic diagram of EPL for EPSP synthase. The EPSPSN-thioester and EPSPSC^{CYS} will form a peptide bond between tyrosine 237 and cysteine 238 by the native chemical ligation. The EPSPSN-thioester (**1**) was expressed as an intein-fusion protein and cleaved by treatment of MESNA as described in *Methods and Materials* whereas EPSPS^{CYS} was prepared from insoluble Met-EPSPS (**3**) by the aid of MAP. CBD, chitin binding domain; MESNA, mercapto-ethanesulphonate.

development since this enzyme since the pathway where this enzyme belongs to does not exist in human.[13] In addition, the appearance of multi-drug resistant (MDR) strains has accelerated the search for the macromolecular targets for new drug discovery.[14,15] Therefore, EPSP synthase can be a good target for the incorporation of various probe using EPL for drug development for the other gram positive pathogens, *Streptococcus pneumoniae* and *Staphylococcus aureus*.[13] We tested general EPL technique for this protein to see if ligation was possible, but failed since the intein-fused C-terminal domain was expressed as an inclusion body and was not refolded properly (Figure 1). Hence different approach was needed to overcome these issues.

Here, we present protein ligation method of EPSP synthase *via* modified version which can be applicable to other large multidomain proteins.

Materials and Methods

Reagents and Enzymes. All chemicals were purchased from Sigma-Aldrich Chemical Co (St. Louis, MO). The wild-type EPSP synthase was purified as described elsewhere.[16,17] IMPACT™-TWIN kit was from New England Biolabs (Beverly, MA). Primers used in this study are listed in Table 1 of supporting information.

Cys-scanning Site-directed Mutagenesis. To determine a possible split residue, site-directed mutagenesis was employed to residues around the 2nd linker such as T236, L238, V239, and E240. A plasmid, pWS250, which encodes the wild-type EPSP synthase was used as a template.[17] These four residues were mutated to cysteine using QuikChange™ Kit (Stratagene) with the *pfx* DNA polymerase as described in manufacturer's manual.

Construction of Plasmids. As illustrated in Figure 1, the thioester tag at the C-terminus of EPSPSN, and an α-cysteine at the N-terminus of EPSPSC are required to facilitate protein ligation.[3,18] These fragments were produced using a pTWIN1 vector (New England Biolabs, Beverly, MA) according to the manufacturer's manual (detailed description in supporting information). The EPSP synthase was split at L238 (see *Results and Discussion*). The resulting plasmids were designated as a pHKN and a pHKC for the **1** and **2'**, respectively (Figure 1). The expression vector for the alternative EPSPSC (Met-EPSPSC^{CYS}, Met-238^{CYS}-427, **3** at Figure 1) was constructed by adding three residues to a plasmid harboring the C-terminal fragment (residues 241-427) using the QuikChange™ Kit (Stratagene) with *pfx* DNA polymerase.[19] To improve the removal of the initiation methionine residue, an extra amount of methionine amino-peptidase (MAP) was produced by a pHK623 plasmid which is compatible with other plasmids used in this study.

Protein Expression and Purification. The cysteine mutants to search for the split site were overexpressed and purified as described elsewhere.[16,17] All the steps were performed at 4 °C, unless otherwise stated. The concentration of protein was determined by the method of Bradford

Table 1. Primers used in this study

Primer	Sequence[a]
T236C	5′-GTC TTA TCA GTC TCC GGG **TTG** **C**TA TTT GGT CGA AGG C-3′
	5′-GCC TTC GAC CAA ATA **GCA** ACC CGG AGA CTG ATA AGA C-3′
L238C	5′-CAG TCT CCG GGT ACT TAT **TGC** GTC GAA GGC GAT GC-3′
	5′-GCA TCG CCT TCG AC**G** **CA**A TAA GTA CCC GGA GAC TG-3′
V239C	5′-CCG GGT ACT TAT TTG **TGC** GAA GGC GAT GCA TCT TCG-3′
	5′-CGA AGA TGC ATC GCC TTC **GCA** CAA ATA AGT ACC CGG-3′
E240C	5′-CCG GGT ACT TAT TTG GTC **TGC** GGC GAT GCA TCT TCG G-3′
	5′-CCG AAG ATG CAT CGC C**GC** **A**GA CCA AAT AAG TAC CCG G-3′
EPSPSN	5′-GGT GGT **TGC** **TCT** **TC**C GCA ATA AGT ACC CGG AGA-3′
EPSPSC	5′-GGT GGT **TGC** **TCT** **TC**C AAC TGC GTC GAA GGC GAT GCA TCT-3′
	5'-GGT **gga** **tcc** tca ggc tgc ctg gct aat-3'
MAP	5′-GAT CGG AAG TCC GGC GCG CT-3′
	5′-GCT GAG GAC GTC GCT TTT ATC CCA CCG ACG GT-3′
EPSPSC^{CYS}	5′-CTT TAA GAA GGA GAT ATA CAT ATG **TGC** GTC GAA GGC GAT GCA TCT TCG-3′
	5′-CGA AGA TGC ATC GCC TTC GAC **GCA** CAT ATG TAT ATC TCC TTC TTA AAG-3′

[a]The sequences for the mutation and restriction site for *Sap* I and *Bam*HI are shown in bold type, while the coding sequences that anneal to the plasmids during the PCR are underlined.

with bovine serum albumin (BSA) as a protein standard.

For the preparation of fragments of EPSPSN-thioester and EPSPSC^{CYS}, plasmids harboring each intein-fused fragment were transformed into the ER2566 strain. Overexpression of the fusion protein was induced by addition of IPTG to a final concentration of 0.4 mM. Then, the cultures were transferred to a 22 °C shaker and incubated for a further 9 hours. The fusion protein was purified by a chitin affinity column. The EPSPSN-thioester was eluted with a buffer containing 50 mM MESNA, while EPSPSC intein-fused protein was expressed as an inclusion body, which was not refolded under the various conditions utilized. Alternatively, the ligation-active EPSPSC was prepared from two different strains. The plasmid encoding Met-EPSPSC (Met-238^{CYS}-427) was transformed into both BL21(DE3) and BL21(DE3) pHK623 strain, which encodes the methionine aminopeptidase (MAP). This construct also expressed inclusion body of EPSPSC but the ligation-active EPSPSC was prepared as described in supporting information and confirmed by MASS-TOF and N-terminal amino acid sequencing.

Expressed Protein Ligation. Ligation reactions were set up as follows. EPSPSN-thioester (**1** at Figure 1) and EPSPSC^{CYS} (**2** at Figure 1) were concentrated to 24 mg/mL (~1 mM) and 51.8 mg/mL (~2 mM), respectively. Two recombinant proteins were mixed in a 1:1 ratio in 500 μl with the final concentration of ~0.5 mM and ~1 mM respectively. 2% (v/v) of thiophenol was added to increase the ligation efficiency by exchanging thiols. The solutions were vigorously vortexed for 30 sec. The reaction was incubated at 37 °C overnight. Ligation efficiencies were determined as described.[20]

Folding and Purification of a Ligated EPSPS. The ligation reaction mixture was diluted in buffer E (50 mM TrisHCl, pH 7.8, 1 mM EDTA and 50 mM NaCl) containing 6 M GdHCl to the final protein concentration of ~ 2 mg/mL. The solution was stirred overnight, centrifuged to remove undissolved particles, and then filtered using a 0.2 μm membrane. The folding of the ligated EPSPS was initiated by diluting the clarified ligation reaction mixture into 40 volumes of buffer E containing 1 mM DTT. The aggregates formed during folding were centrifuged and removed by the 0.2 μm filter membrane. The folded proteins were concentrated and loaded onto the FPLC MonoQ column equilibrated with 50 mM TrisHCl, pH 7.8, 1 mM EDTA, 1 mM DTT. The proteins were eluted with 0-0.5 M NaCl linear gradient. The ligated EPSP synthase was eluted at ~0.4 M NaCl, while the unligated EPSPSN eluted at ~0.3 M NaCl.

Circular Dichroism and Enzyme Activity Assay. The far-UV CD spectra of various EPSP synthase were measured using an AVIV 202SF spectrophotometer (AVIV biomedical Inc., NJ, USA) at 25 °C and a 0.2 cm-pathlength quartz cell. The concentration of the proteins was 10 μM in 20 mM potassium phosphate buffer, pH 7.5, 1 mM EDTA, and 1 mM DTT. Background corrections were made in all spectra. The enzyme activity was measured routinely by the reverse coupled assay of Lewendon and Coggins.[21] All the enzyme assays were performed at 25 °C.

Results

L238 as a Split Site. EPL uses the chemistry of the native chemical ligation to ligate recombinantly expressed proteins.[5,22] For the native chemical ligation to work, there are two requirements: firstly a peptide or protein containing an α-thioester, and secondly a peptide or protein containing an N-terminal cysteine.[2,3,22] The N-terminal cysteine residue at the second peptide or protein (**2** at Figure 1) limits the application of EPL even though a few different approaches have been designed.[3] In the case of *E. coli* EPSP synthase, the C-terminal domain is discontinuous in the primary

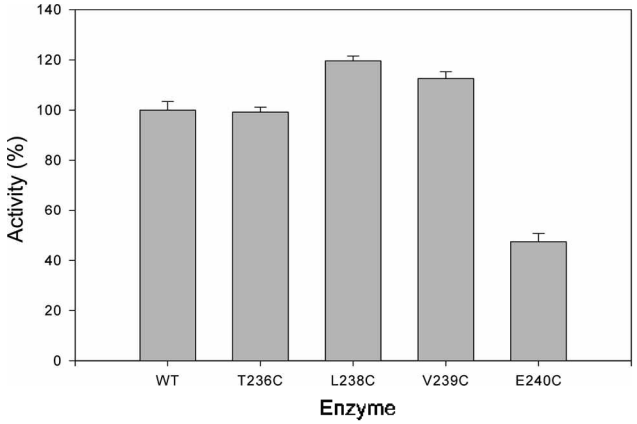

Figure 2. Specific activity of mutant enzymes as a percentage of wild-type activity.

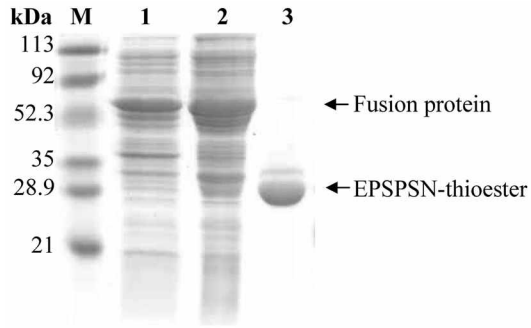

Figure 3. Purification of EPSPSN-thioester analyzed on 12% SDS-PAGE. Lane M: molecular weight marker (BioRad), lane 1: uninduced cell, lane 2: induced cell, lane 3: EPSPSN after cleavage. Fifty-nine kDa of the fusion protein (the N-terminal fragment, 25 kDa; Mxe GyrA, 28 kDa; CBD, ~6 kDa) was expressed and purified by chitin bead. Then EPSPSN-thioester was cleaved by the thiol reagent MESNA.

structure (Figure 1A).[23] The residues in the second linker, where there are no cysteine residues, were screened for possible split sites by introducing a cysteine residue using site-directed mutagenesis. The introduction of a cysteine residue for protein ligation may affect not only the function but also the structure of the protein, so the activity assay of the cysteine mutant EPSP synthase proteins at the second linker region is, therefore, a prerequisite prior to splitting the protein. T236, L238, V239, and E240 at the second linker region of EPSP synthase were mutated to cysteine as described above. These residues were chosen for various reasons. In an independent work by Chen *et al.*, T236 was proved to be tolerable to the insertion of extra residues.[24] L238 was preceded by a tyrosine residue which is known to increase the thiolysis of intein fused proteins (Figure 1).[18,20] The side chain of V239 has a similar bulkiness to that of a cysteine. E240 was the original splitting site for the divide-and-conquer strategy we used to study this protein using NMR.[19,25] The result of the activity assay of these mutants was shown in Figure 2. Except E240C, all mutant protein showed similar or greater activity than that of wild-type. Because none of these residues are involved in either substrate or inhibitor binding, it may stand to reason that these mutations may not change the activity. According to the ternary complex (EPSPS·S3P·GLP) structure of Schönbrunn *et al.*, E240 forms two additional hydrogen bonds with M270 and Q271 upon substrate binding, which may play a role in stabilizing the domain closure.[13] However, the mutated cysteine may not form the bonds due to its shorter side chain. This may account for the loss of activity of E240C to ~50%. Of these residues, L238 was chosen as a split site, since mutation to cysteine is tolerable and the preceding tyrosine (Y237) is preferred for the production of an α-thioester-tagged protein.[20]

Preparation of the Thioester-tagged N-Terminal Fragment (EPSPSN-thioester). The construction of the expression vector for the production of EPSPSN-thioester was described above. The fusion protein was expressed as a soluble protein comprising up to ~30% of total cell protein. The EPSPSN-thioester was purified by one-step chitin affinity chromatography (Figure 3) and 110.4 mg/mL was obtained from 4 L LB culture.

Preparation of the α-Cysteine C-Terminal Fragment (EPSPSC^{CYS}). The intein-fused construct for EPSPSC^{CYS} encoding residues Met^{237}-238^{CYS}-427 was overexpressed as an inclusion body. As described elsewhere,[19] the deletion of the first 20 residues in either the full-length EPSP synthase

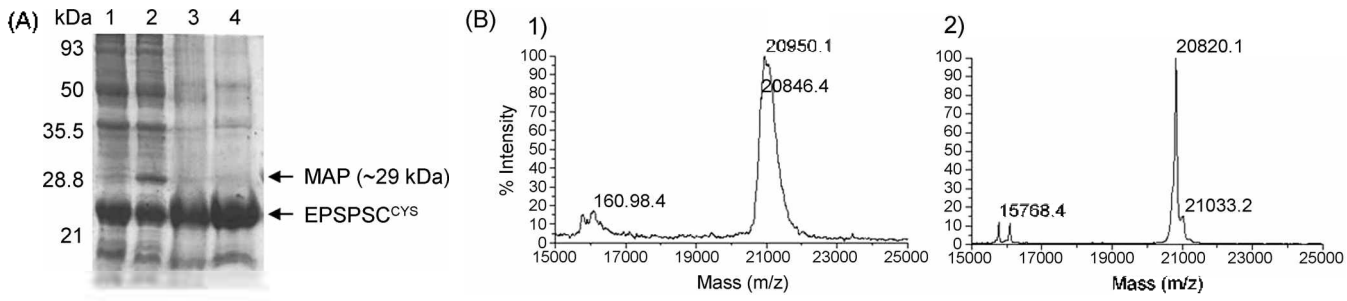

Figure 4. Purification of EPSPSC^{CYS}. (A) 12.5% SDS-PAGE of EPSPSC^{CYS}, one (1,3) prepared from *E. coli* strain, the other (2,4) prepared from the MAP coexpressing strain: lane 1 and 2, overexpression of EPSPSC^{CYS}, and lane 3 and 4, purification of inclusion bodies from 1 and 2 respectively. In lane 2, the expression of MAP was also detected at ~29 kDa. (B) Mass spectrometry data of EPSPSC^{CYS} prepared from two strains. Spectra 1) and 2) represent the proteins from lane 3 and 4 at SDS-PAGE, respectively. It clearly demonstrated that EPSPSC^{CYS} from lane 4 showed a sharper peak with a value close to the expected molecular weight of methionine cleaved EPSPSC^{CYS}. The calculated molecular weights of EPSPSC^{CYS} and Met-EPSPSC^{CYS} are 20,818.76 Da and 20,949.95 Da, respectively.

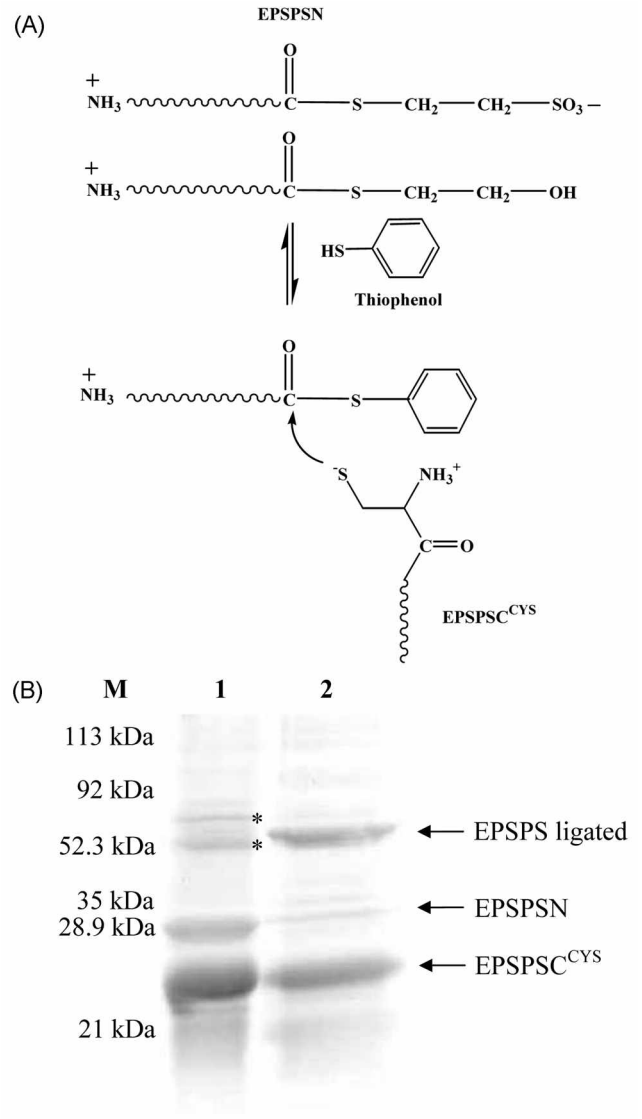

Figure 5. Schematic representation of thiol exchange reaction (A) and SDS-PAGE analysis of protein ligation (B). Possibly two kinds of thioester-tagged protein exist in the protein ligation mixture since β-mercaptoethanol was carried over from the dissolving buffer for EPSPSC^{CYS} to the ligation mixture. In A, thiophenol added in reaction will exchange thiols with less active thiol groups. (B) SDS-PAGE shows the ligation reaction in the presence of 2% thiophenol. Lane 1 is prior to ligation reaction, and lane 2 is post ligation reaction. As can be seen, most of the EPSPSN-thioester disappeared after the ligation and ligated EPSPS appeared with the corresponding size. The asterisks are minor contaminants from the EPSPSC^{CYS} preparation.

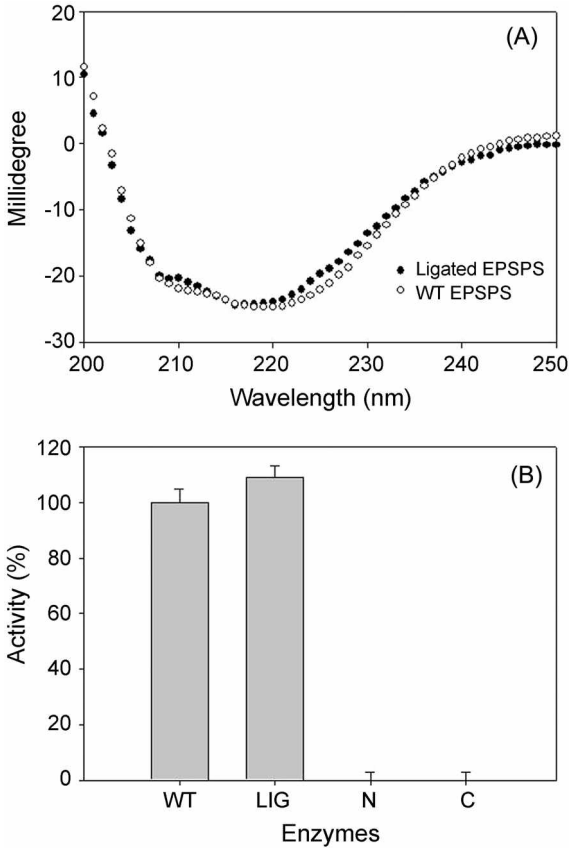

Figure 6. CD (A) and activity assay (B). (A) The far-UV CD spectra (200-250 nm) of wild-type (open circle) and ligated EPSP synthase (filled circle) were displayed. (B) The specific activities were plotted as percentage of wild-type activity. N and C are EPSPSN α-thioester and EPSPSC^{CYS} respectively.

or its isolated C-terminal domain led to the formation of inclusion bodies. It was, therefore, not surprising that the intein-fused EPSPSC was expressed as an inclusion body and was unable to fold correctly *in vitro* (data not shown). To obtain the EPSPSC^{CYS}, which is essential for the expressed protein ligation, a new method was adopted. During translation or after the completion of translation in *E. coli* the N-formyl moiety and the methionyl residue are cleaved off in a significant fraction of the cytosolic proteins by deformylase and methionine aminopeptidase (MAP).[26,27] Although the endogenous MAP might be sufficient for the cleavage of the initiation methionine, its coexpression will help cleave the N-terminal formyl methionine residue more efficiently.[28-30] Therefore, the gene encoding MAP was amplified and inserted clockwise into the pACYC184 for plasmid compatibility. The EPSPSC was overexpressed with or without pHK623 harboring MAP. The EPSPSC^{CYS} protein was prepared from inclusion body (Figure 4A) and analyzed by MALDI-TOF, showing that EPSPSC^{CYS} coexpressed with pHK623 had most of the initiation methionine removed, while the protein expressed alone showed a broad peak, which means only some of the methionine residues were removed (Figure 4B). Additionally, N-terminal amino acid sequencing also confirmed the MOLDI-TOF data.

Protein Ligation. For the protein ligation, 100 μL of 1 mM of EPSPSN-thioester (final 0.5 mM) was mixed with 100 μL of ~2 mM of EPSPSC^{CYS} (final 1 mM). Ligation was performed in the presence of ~4 M GdHCl as the EPSPSC^{CYS} precipitated in the absence of guanidinium. Chaotropic agents, such as guanidium chloride and urea, have been shown not to inhibit the protein ligation.[3,4,20] Thiophenol, as a thiol additive in the native chemical ligation, has been reported to generate a more reactive phenyl thioester *via* thiol exchange and keep the N-terminal

cysteine residue of a second peptide reduced (Figure 5A).[31] The initial protein ligation was ineffective without the addition of thiophenol, presumably due to β-mercaptoethanol from the buffer used to dissolve α-cysteine EPSPSC^{CYS} (data not shown). β-mercaptoethanol may undergo thiol exchange with the ethanesulfonic ester of the EPSPSN and make it less ligation-active.[20] DTT and β-mercaptoethanol are known to less activate the protein ligation due to its low leaving group ability compared with thiophenol and MESNA.[20] As anticipated, the addition of thiophenol to a final concentration of 2% (~180 mM) drastically increased the ligation efficiency up to ~85% (Figure 5B). It also demonstrated that the ligation of EPSPSN-thioester with the EPSPSC^{CYS} from a strain without extra MAP coexpressed was less effective than with that from the strain coexpressing MAP, since many of EPSPSC^{CYS} still have an initiation methionine residue which blocks the protein ligation (data not shown).

Folding and Purification of Ligated EPSP Synthase. The ligation reaction was stopped by diluting the final protein concentration of ~2 mg/mL in a buffer containing 6 M GdHCl. The clarified solution contained unreacted EPSPSN-thioester and EPSPSC^{CYS}, and the ligated EPSP synthase. Refolding of the ligated product was initiated by twenty fold dilution. As described elsewhere,[19] the EPSPSC^{CYS} was unable to fold properly, so that during the refolding step, most of the EPSPSC^{CYS} aggregated and precipitated, hence making the subsequent purification step easier. The soluble fraction was further purified by MonoQ chromatography. The ligated products were separated from EPSPSN-thioester, although some of them were eluted in the same area of the gradient as EPSPSN-thioester. The purity of the eluted protein from the MonoQ was ~95% pure as judged by coomassie stained 10% SDS-PAGE.

CD and the Enzyme Activity Assay. CD was measured for both the wild-type EPSP synthase and ligated EPSP synthase for comparison. As illustrated in Figure 6A, the spectra showed that both proteins have distinct secondary structure with the lowest ellipticities at 210 and 220 nm. The CDPro analysis revealed that the secondary structure content for ligated EPSP synthase is similar to that for wild-type EPSP synthase. The activity of the EPSP synthase was measured by the reverse coupled assay of Lewendon and Coggins.[21] The activity of ligated EPSP synthase was slightly higher than that of wild-type enzyme. However, none of the EPSPSN-thioester and EPSPSC^{CYS} alone showed the enzymatic activity (Figure 6B).

Discussion

Due to the limitations of solid phase peptide synthesis (SPPS), native chemical ligation is restricted by the size of the protein.[4] Recombinant preparation of the thioester-tagged proteins from the intein fusion protein has drastically expanded the application of native chemical ligation,[18,20] so that the number and size of proteins engineered using EPL are continuously increasing.[2] Although EPL can be theoretically applied to any size of proteins, the technique has only been used for the ligation of a segment or a domain comprising up to 50 to 100 amino acid residues within a large piece of a protein.[32]

In this work, the extension of EPL was demonstrated: ligation of two domains, ~25 and 20 kDa respectively. The application of EPL to a larger protein may not always be easy since larger proteins are often composed of multi domains. In addition, not all the domains can be isolated stably.[33,34] In the case of EPSP synthase, the topology of this enzyme is somewhat unique and the preparation of the EPSPC^{CYS} from the intein construct was impossible. There are a few possible ways of obtaining EPSPC^{CYS}. For instance, this protein could be fused to other proteins such as GST with an enterokinase cleavage site in the linker region, or fused to the end of ubiquitin which can be cleaved directly after ubiquitin by ubiquitinase. However these alternative fusion approaches are only applicable under circumstance where the fusion proteins are expressed solubly or at least can be refolded. This is not the case for EPSPSC. Alternative to these, the N-terminal processing of the protein in *E. coli* was considered. Many of the initiation N-formyl methionines are cleaved off during or after translation by MAP, and its removal is dependent on the penultimate residue.[26,27,35] In addition, it is also reported that initiation methionine can be almost completely removed by endogenous MAP.[35] Very recently, independent but concurrent to this idea, Camareo *et al.* (2001) used the endogenous MAP to generate a cyclized protein *in vivo* using an intein.[36] Although the endogenous MAP may be enough to process the initiation methionine, coexpression of this protein with the target protein increases the removal of the methionine residue.[26-28,30] As shown in Figure 4B, the mass spectrometry data clearly demonstrated that the EPSPSC^{CYS} coexpressed with MAP showed higher homogeneity than the one expressed in its absence. The subsequent ligation reactions with these proteins indicated that the amount of the ligated product is dependent on the concentration of the α-cysteine residue. This work clearly demonstrated that the second peptide or protein, soluble or not, with α-cysteine at its N-terminus can be produced homogeneously by this approach.

The rate of ligation is dependent on the leaving group ability of α-thioester in a peptide or a protein. In our case, the efficiency of ligation of two fragments was increased up to ~85% by adding thiophenol, which functions as a better leaving group. In the reaction mixture, three thiol reagents: β-mercaptoethanol, MESNA, and thiophenol compete with one another (Figure 5A). In the EPSPSN preparation, ~50 mM MESNA was used to produce the EPSPSN-thioester, this has been reported to be as effective as thiophenol in protein ligation with the advantage of being odorless.[20] The ligation reaction without any thiol additives showed less ligated product implying that MESNA may not compete out the 20 mM β-mercaptoethanol which was carried over from the EPSPSC^{CYS} preparation and may form unproductive thioester (data not shown). The addition of 2% (~180 mM) thiophenol greatly increased the ligation efficiency, due to

the reversal of unproductive thioester to productive one *via* thiol exchange.[31] The addition of an excess of MESNA also increased the ligation but to a lesser degree (data not shown).

The ligated product in the native state may not require refolding,[32] but in the denatured state should proceed to refolding to obtain an active enzyme.[4,5,20] The preliminary data of the EPSP synthase refolding experiment showed that rapid dilution of the denatured EPSP synthase recovered around 80% of the wild-type activity. However in a large scale of protein folding, the folding of this protein was inefficient due to unique topology of the protein. The elution profile of the refolded proteins on MonoQ chromatography showed that the EPSPSN-thioester was eluted around 0.3 M NaCl with a single peak but the ligated EPSP synthase eluted around 0.4 M NaCl with inseparable multiple peaks (data not shown). These multiple peaks suggest that there may be heterogeneities in the protein sample. The analysis of CD spectra showed that the wild-type and ligated EPSP synthase have a very similar content of secondary structure. Additionally, the specific activity data revealed that the ligated protein showed similar activity to that of L238C mutants, meaning the refolding of the protein in our scale was successful. This work demonstrated that incorporation of various optical probes *via* ligation method presented in this study is possible to large proteins, especially whose second domain or fragment is expressed insolubly.[4,20,32,37] It is also extended to segmental stable isotope labeling for NMR. For example, membrane proteins which are hard to express in *E. coli* and are often expressed as insoluble proteins, may be well suited to this method combined with other peptide and protein chemistry.[6,38-40]

Acknowledgement. This work is dedicated to the late Prof. Jeremy Evans for his scientific guidance and support. This work was supported by the Korea Polar Research Institute (Grant PE07060 to H. J. K. and S.-H. K.) and by the Ministry of Maritime Affairs and Fisheries (Grant PM44400 to H. J. S.).

References

1. Muralidharan, V.; Muir, T. W. *Nat. Methods* **2006**, *3*, 429.
2. Hofmann, R. M.; Muir, T. W. *Curr. Opin. Biotechnol.* **2002**, *13*, 297.
3. Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249.
4. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776.
5. Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 388.
6. Lu, W.; Gong, D.; Bar-Sagi, D.; Cole, P. A. *Mol. Cell.* **2001**, *8*, 759.
7. Shi, J.; Muir, T. W. *J. Am. Chem. Soc.* **2005**, *127*, 6198.
8. Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. *Proteins* **2001**, *42*, 38.
9. Kishore, G. M.; Shah, D. M. *Annu. Rev. Biochem.* **1998**, *57*, 627.
10. Haslam, E. *Shikimic Acid: Metabolism and Metabolites*; John Wiley: Chichester, 1993.
11. Steinrucken, H. C.; Amrhein, N. *Biochem. Biophys. Res. Commun.* **1980**, *94*, 1207.
12. Franz, J.; Mao, M. K.; Sikorski, J. A. *Glyphosate: A Unique Global Herbicide*; Oxford University Press: New York, 1997.
13. Schonbrunn, E.; Eschenburg, S.; Shuttleworth, W. A.; Schloss, J. V.; Amrhein, N.; Evans, J. N.; Kabsch, W. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1376.
14. Dessen, A.; Di Guilmi, A. M.; Vernet, T.; Dideberg, O. *Curr. Drug Targets Infect. Disord.* **2001**, *1*, 63.
15. Coggins, J. R.; Abell, C.; Evans, L. B.; Frederickson, M.; Robinson, D. A.; Roszak, A. W.; Lapthorn, A. P. *Biochem. Soc. Trans.* **2003**, *31*, 548.
16. Shuttleworth, W. A.; Hough, C. D.; Bertrand, K. P.; Evans, J. N. S. *Protein Engineering* **1992**, *5*, 461.
17. Shuttleworth, W. A.; Evans, J. N. S. *Biochemistry* **1994**, *33*, 7062.
18. Evans, T. C., Jr.; Benner, J.; Xu, M. Q. *J. Biol. Chem.* **1999**, *274*, 3923.
19. Kim, H. J.; Kim, H. W.; Kang, S.-H. *J. Microbiol. Biotechnol.* **2007**, *17*, 1385.
20. Evans, T. C., Jr.; Benner, J.; Xu, M. Q. *Protein Sci.* **1998**, *7*, 2256.
21. Lewendon, A.; Coggins, J. R. *Biochem. J.* **1983**, *213*, 187.
22. Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.
23. Stallings, W. C.; Abdelmeguid, S. S.; Lim, L. W.; Shieh, H. S.; Dayringer, H. E.; Leimgruber, N. K.; Stegeman, R. A.; Anderson, K. S.; Sikorski, J. A.; Padgette, S. R.; Kishore, G. M. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5046.
24. Chen, L.; Pradhan, S.; Evans, T. C., Jr. *Gene* **2001**, *263*, 39.
25. Stauffer, M. E.; Young, J. K.; Helms, G. L.; Evans, J. N. *J. Biomol. NMR* **2001**, *20*, 387.
26. Gonzales, T.; Robert-Baudouy, J. *FEMS Microbiol. Rev.* **1996**, *18*, 319.
27. Bradshaw, L. D.; Padgette, S. R.; Kimball, S. L.; Wells, B. H. *Weed Technology* **1997**, *11*, 189.
28. Van Valkenburgh, H. A.; Kahn, R. A. *Methods Enzymol.* **2002**, *344*, 186.
29. Adachi, K.; Yamaguchi, T.; Yang, Y.; Konitzer, P. T.; Pang, J.; Reddy, K. S.; Ivanova, M.; Ferrone, F.; Surrey, S. *Protein Expr. Purif.* **2000**, *20*, 37.
30. Sandman, K.; Grayling, R. A.; Reeve, J. N. *Biotechnology* **1995**, *13*, 504.
31. Muir, T. W.; Dawson, P. E.; Kent, S. B. *Methods Enzymol.* **1997**, *289*, 266.
32. Camarero, J. A.; Shekhtman, A.; Campbell, E. A.; Chlenov, M.; Gruber, T. M.; Bryant, D. A.; Darst, S. A.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8536.
33. Minard, P.; Hall, L.; Betton, J. M.; Missiakas, D.; Yon, J. M. *Protein Eng.* **1989**, *3*, 55.
34. Sherman, M. A.; Chen, Y.; Mas, M. T. *Protein Sci.* **1997**, *6*, 882.
35. Lathrop, B. K.; Burack, W. R.; Biltonen, R. L.; Rule, G. S. *Protein Expr. Purif.* **1992**, *3*, 512.
36. Camarero, J. A.; Fushman, D.; Cowburn, D.; Muir, T. W. *Bioorg. Med. Chem.* **2001**, *9*, 2479.
37. Camarero, J. A.; Cotton, G. J.; Adeva, A.; Muir, T. W. *J. Pept. Res.* **1998**, *51*, 303.
38. Valiyaveetil, F. I.; MacKinnon, R.; Muir, T. W. *J. Am. Chem. Soc.* **2002**, *124*, 9113.
39. Miller, J. S.; Dudkin, V. Y.; Lyon, G. J.; Muir, T. W.; Danishefsky, S. J. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 431.
40. Flavell, R. R.; Huse, M.; Goger, M.; Trester-Zedlitz, M.; Kuriyan, J.; Muir, T. W. *Org. Lett.* **2002**, *4*, 165.