The in Vitro Ligation of Bacterially Expressed Proteins Using an Intein from Methanobacterium thermoautotrophicum*

(Received for publication, November 23, 1998, and in revised form, December 18, 1998)

Thomas C. Evans, Jr., Jack Benner, and Ming-Qun Xu
From New England Biolabs, Inc., Beverly, Massachusetts 01915-5599

The smallest known intein, found in the ribonucleoside diphosphate reductase gene of Methanobacterium thermoautotrophicum (Mth RIR1 intein), was found to splice poorly in Escherichia coli with the naturally occurring proline residue adjacent to the N-terminal cysteine of the intein. Splicing proficiency increased when this proline was replaced with an alanine residue. However, constructs that displayed efficient N- and C-terminal cleavage were created by replacing either the C-terminal asparagine or N-terminal cysteine of the intein, respectively, with an alanine. Furthermore, these constructs were used to specifically generate complementary reactive groups on protein sequences for use in ligation reactions. Reaction between an intein-generated C-terminal thioester on E. coli maltose-binding protein (43 kDa) and an intein-generated cysteine at the N terminus of either T4 DNA ligase (56 kDa) or thioredoxin (12 kDa) resulted in the ligation of the proteins through a native peptide bond. Thus the smallest of the known inteins is capable of splicing and its unique properties extend the utility of intein-mediated protein ligation to include the in vitro fusion of large, bacterially expressed proteins.

Inteins (1), the protein equivalent of the self-splicing RNA introns, catalyze their own excision from a precursor protein with the concomitant fusion of the flanking protein sequences, known as exteins (reviewed in Refs. 2–4). Almost 100 inteins have been identified (5) and can be grouped into three classes: 1) the inteins containing a homing endonuclease between the two splicing domains, 2) the mini-inteins, which lack the homing endonuclease, and 3) a newly described trans-splicing intein, which catalyzes its own excision from a precursor protein with a C-terminal thioester (9). This intein may be close to the minimum amino acid sequence of the known inteins is capable of splicing and its unique properties extend the utility of intein-mediated protein ligation to include the in vitro fusion of large, bacterially expressed proteins.

Studies into the mechanism of splicing led to the development of a protein purification system that utilized thiol-induced cleavage of the peptide bond at the N terminus of the intein (9). Purification with this system generated a bacterially expressed protein with a C-terminal thioester (9). Two research groups then applied the chemistry described for native chemical ligation (10) to fuse a synthetic peptide with an N-terminal cysteine to a bacterially expressed protein possessing a C-terminal thioester (11, 12). This technique, known as intein-mediated protein ligation (IPL)² or also as expressed protein ligation, represented an important advance in protein semi-synthetic techniques (reviewed in Refs. 13 and 14). However, the generality of IPL was limited by the use of a synthetic peptide as a ligation partner.

We describe the next major advance in intein-mediated protein ligation, which is the modulation of the Mth RIR1 intein for the facile isolation of a protein with an N-terminal cysteine for use in the in vitro fusion of two bacterially expressed proteins. Furthermore, the Mth RIR1 mini-intein, the smallest known protein splicing element, was found to be capable of splicing. These results significantly expand the utility of IPL to include the labeling of extensive portions of a protein for NMR analysis and the isolation of a greater variety of cytotoxic proteins. In addition, this advance opens the possibility of labeling the central portion of a protein by ligating three fragments in succession.

EXPERIMENTAL PROCEDURES

Mth RIR1 Synthetic Gene Construction—The gene encoding the Mth RIR1 intein along with native N- and C-extein residues (Fig. 1; Ref. 7) was constructed using 10 oligonucleotides (New England Biolabs, Beverly, MA) comprising both strands of the gene and over-lapping at about 20 base pairs. 1) 5’-TGACGAGCAACACAACTCCGTCATTGCCGTCAGC-3’; 2) 5’-GGGCGA-3’; 3) 5’-TGATTGCGCGCTCGGTATCCATGCCCCTCTAGGTATTCTTCTTCC-3’; 4) 5’-GGGCGA-3’; 5) 5’-pCCAGCGCCACGCAGGCCACGGTGCAG-3’; 6) 5’-pTGATTCGCGGCTCTGGCTACCCATGGCCCCTCAGGTTTCTTCC-3’.

Of the mini-inteins, the smallest is the 134-amino acid intein found in the ribonucleoside diphosphate reductase gene of Methanobacterium thermoautotrophicum (Mth RIR1 intein; Ref. 7). This intein may be close to the minimum amino acid sequence needed to promote splicing, and interestingly, it has a proline residue N-terminal to the first amino acid of the intein, Pro–1 (see Fig. 1), which was shown to inhibit splicing in an intein found in the 69-kDa vacular ATPase subunit of Saccharomyces cerevisiae (See Vma intein; Ref. 8).

The abbreviations used are: IPL, intein-mediated protein ligation; MESNA, the sodium salt of 2-mercaptoethanesulfonic acid; MBF, maltose-binding protein; CBD, chitin-binding domain; N-B, a fusion chitin-binding domain; M-R-B, a fusion protein consisting of maltose-binding protein-Mth RIR1 intein-chitin-binding domain; IPTG, isopropyl-β-D-thiogalactopyranoside; PAGE, poly-acrylamide gel electrophoresis.
incorporates 61 silent base mutations in 48 of the 134 codons. The 
oligonucleotides were annealed by mixing at equimolar ratios (200 ng) in a 
ligation buffer (50 mM Tris- HCl, pH 7.5, 10 mM dithiothreitol, 1 mM ATP, and 25 
µg of bovine serum albumin) followed by heating to 95 °C. After cooling to room temperature, the 
annealed and ligated oligonucleotides were inserted into the XhoI and 
AgeI sites of pMYB5 (New England Biolabs), replacing the 
Sce VMA increate and the plasmid pMRB8P. 

Mutagenesis of the Mth RIR1 Intein—The unique XhoI and SpeI sites flanking the N-terminal splice junction and the unique 
BglII and AgeI sites flanking the C-terminal splice junction allowed substitution of 
amino acid residues by linker replacement. Pro-1, the proline residue 
preceding the intein in pMRB8P, was substituted with alanine or glycine 
to yield pMRB8A and pMRB8G1, respectively. Substitution of Pro 
1-Cys 2 with Gly-Ser or Gly-Ala yielded pMRB9GS and pMRB9GA, respectively. Replacing Asn 
3 with Ala in pMRB8G1 resulted in pMRB10G. The following links were used for substitution of 
the native amino acids at the splice junctions. Each linker was formed by 
annealing two synthetic oligonucleotides as described above. 

Protein Purification with the C-terminal Cleavage Construct—Protein 
purification was performed as described above with buffer A replaced 
buffer C 20 mM Tris- HCl, pH 8.5, containing 500 mM NaCl and 
buffer B replaced by buffer D 20 mM Tris- HCl, pH 7.0, containing 
500 mM NaCl). Also, following equilibration of the column in buffer D 
the cleavage reaction proceeded overnight at room temperature. Protein 
cleavage products were determined using the Bio-Rad protein assay. 

Protein-Protein Ligation Using IEL—Freshly isolated thiostrept 
tagged protein was mixed with freshly isolated protein containing an 
N-terminal cysteine residue (starting concentration, 1–200 µM). The solution 
was concentrated to a Centriprep 3 or Centricon 30 apparatus (Millipore Corporation, Bedford, MA) then with a Centricon 3 or 
Centricon 10 apparatus to a final concentration of 0.15–1.2 mM for each 
protein. Ligation reactions proceeded overnight at 4 °C and were visualized 
using SDS-PAGE with 12% Tris-glycine gels (Novex Experimental 
Technology, San Diego, CA) stained with Coomassie Brilliant Blue. 

Factor Xa Cleavage of MBP-T4 Ligase Fusion Protein and Protein 
Sequencing—2 mg of ligation reaction involving MBP and T4 DNA 
ligase was bound to 3 ml of amylase resin (New England Biolabs) 
equilibrated in buffer A (see above). Unreacted T4 DNA ligase 
was rinsed from the column with 10 column volumes of buffer A. Unligated 
MBP and the MBP-T4 DNA ligase fusion protein were eluted from the 
amylase resin using buffer E (20 mM Tris- HCl, pH 7.5, containing 500 
mM NaCl and 10 mM maltose). Overnight incubation of the eluted 
protein with a 200:1 protein:bovine factor Xa (New England Biolabs) 
was performed on a Precise 494 protein sequencer (PE Applied Biosystems, Foster City, CA). 

RESULTS 

Splicing and Cleavage Activity of the Mth RIR1 Intein—The 
splicing activity of the Mth RIR1 intein with its 5 native N- and 
C-extein residues was investigated by expressing it as an 
in-frame fusion between E. coli maltose-binding protein (15) and the 
chitin-binding domain (16) from Bacillus circulans. In this 
protein context splicing products were detected (Fig. 2, lane 1), although 
the majority of the protein remained in the precursor form (M-R-B). Splicing efficiency was increased by replacing 
Pro-1 to Ala (Fig. 2, lane 3). Further, the Pro-1 to Ala or 
Pro-2 to Gly mutants also displayed cleavage at the N- 
and C-terminal junctions of the intein (Fig. 2, lanes 3 and 5). 
The identity of splicing and cleavage products were confirmed by Western blot analysis using anti-MBP and anti-CBD 
polyclonal antibodies (data not shown). 

The cleavage and/or splicing activity of the M-R-B precursor 

3 R. Chong, unpublished data.
was more proficient when protein synthesis was induced at 15 °C than when the induction temperature was raised to 37 °C (Fig. 2). Replacement of Pro
\(^{-1}\) with a Gly and Cys
\(^{-1}\) with a Ser resulted in a double mutant, M-R-B (Pro
\(^{-1}\) → Gly/Cys
\(^{-1}\) → Ser), which showed only in vivo C-terminal cleavage activity when protein synthesis was induced at 15 °C but not at 37 °C (Fig. 2, lanes 7 and 8). Another double mutant, M-R-B (Pro
\(^{-1}\) → Gly/Cys
\(^{-1}\) → Ala) displayed slow cleavage, even at 15 °C, which allowed the accumulation of substantial amounts of the precursor protein (data not shown) and showed potential for use as a C-terminal cleavage construct for protein purification.

**Purification Using C- and N-terminal Cleavage Activity—**

The C- and N-terminal cleavage constructs of the Mth RIR1 intein were used to purify T4 DNA ligase or thioredoxin with an N-terminal cysteine. Amsterdam cleavage activity could not be used because it would prevent ligation. Protein fusion occurred at 20–40% efficiency when the induction temperature was raised to 37 °C (Fig. 2, lanes 7 and 8). The full-length fusion protein consisting of MBP-intein(N)-CBD was separated from cell extract by binding the CBD portion of the protein to a chitin resin. Overnight incubation in the presence of 100 mM MESNA induced ligation of the peptide bond prior to the N terminus of the intein and created a thioester on the C terminus of MBP. The C-terminal cleavage vector, intein(C), had the Pro
\(^{-1}\) → Gly/Cys
\(^{-1}\) → Ala double mutation. The precursor CBD-intein(C)-T4 DNA ligase was isolated from induced E. coli cell extract by binding to a chitin resin as described for N-terminal cleavage. Fission of the peptide bond following the C-terminal residue of the intein resulted in the production of T4 DNA ligase with an N-terminal cysteine. Ligation occurred when the proteins containing the complementary reactive groups were mixed and concentrated, resulting in a native peptide bond between the two reacting species.

**Discussion**

The C-terminal cleavage activity of the mutated Mth RIR1 intein advanced IPL technology by providing a means to isolate proteins possessing an N-terminal cysteine to act as substrates in the in vitro fusion of large, bacterially expressed proteins. Initially, an intein that cleaves in vivo was tested for the ability to generate a protein with an N-terminal cysteine. However, the side chain of the N-terminal cysteine residue appeared to be modified in vitro by an unidentified pathway (data not shown). Although this problem could be circumvented using a protease to cut on the N-terminal side of a cysteine residue, concern over nonspecific proteolysis and the need to remove the protease after cleavage limited its usefulness. Interestingly, C-terminal cleavage using the Mth RIR1 intein appeared to protect the cysteine residue until it could be released in vitro. A recently developed Sec VMA intein with thiol-inducible C-terminal cleavage activity could not be used because it would undergo splitting instead of cleavage with an N-terminal cysteine on the target protein (18).

The concentration dependence of the ligation reaction was probably due to the need to increase the ligation reaction rate to effectively compete with thioester hydrolysis, which would prevent ligation. Protein fusion occurred at 20–40% efficiency at 6.5–8.5 mg/ml of each reactant (data not shown), although
greater extents of reaction (50–60%, Fig. 3) were observed at higher protein concentrations. Many proteins can exist in solution at the lower concentrations, indicating that IPL will be useful for a wide range of applications. However, these conditions are problematic for some proteins, and future work may determine procedures that will lower this concentration requirement.

N-terminal amino acid sequencing through the ligation junction demonstrated that the two proteins were fused tail-to-head in a continuous polypeptide chain and had not fused to form an unusual branched structure. Furthermore, these data reinforce past studies reporting that a native peptide bond is formed using native chemical ligation chemistry (10) because the polypeptide sequencing reaction requires a peptide bond between amino acid residues.

Previously, studies with the Sce VMA intein reported that splicing was inhibited when a proline replaced the naturally occurring glycine at the –1 position (8). However, the Mth RIR1 intein has a naturally occurring proline at this position and was thought to be able to splice with this unique amino acid. The low splicing activity of the Mth RIR1 intein shows that it is capable of splicing but that it may not be folding properly when expressed in E. coli. Alternatively, this intein may require more native extein sequence than provided or require a cofactor such as a prolyl isomerase to promote proficient splicing activity.

The Mth RIR1 intein primary sequence was compared with the amino acid sequence and crystal structure of another mini-intein, the Mxe GyrA intein (19, 20). Most of the amino acids that form two α-helices and a disordered region in the Mxe GyrA intein appeared to be missing in the Mth RIR1 intein. The α-helical and disordered regions were previously found not to be required for splicing of the Ssp DnaB intein (21), and this portion of the protein may only serve as a linker. The small size of this region in the Mth RIR1 intein may decrease its stability and may account for some of its induction temperature-dependent activity.

The mechanism of the induction temperature-dependent splicing and cleavage activity has yet to be determined, but it may be due to reactions occurring at the C terminus of the intein. C-terminal cleavage was more severely affected by induction temperature than N-terminal cleavage activity (Fig. 2). It is also possible that the Mth RIR1 intein could be misfolding in E. coli when induced at the higher temperature, an interesting possibility considering that M. thermoautotrophicum is a thermophilic bacteria.

In conclusion, this report demonstrated that the smallest known intein, the Mth RIR1 intein, along with its 5 native extein residues was capable of splicing. Furthermore, this intein was capable of generating both thioester-tagged proteins and proteins with an N-terminal cysteine. The latter was of particular importance because it facilitated the next major advance in intein-mediated protein ligation, which is the fusion of two large, bacterially expressed proteins. This paves the way for greater freedom in the labeling of proteins for NMR analysis, the isolation of cytotoxic proteins, and in the future the controlled fusion of three bacterially expressed proteins.

**Acknowledgments**—We thank Bill Jack, Inca Ghosh, Francine Perler, Eric Adam, Lixin Chen, Maurice Southworth, Shaorong Chong, Eric Clintor, Chudi Guan, Richard Whitaker, Marilena Hall, and Fana Mersha for valuable discussions and assistance; Shaorong Chong for the gift of the pBYT4 plasmid; Eric Adam and Sanjay Kumar for assistance in amino acid alignments of the Mth RIR1 intein; and Don Comb for support and encouragement.

**REFERENCES**

1. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J., and Belfort, M. (1994) *Nucleic Acids Res.* **22**, 1125–1127
2. Perler, F. B., Xu, M.-Q., and Paulus, H. (1997) *Curr. Opin. Chem. Biol.* **1**, 292–299
3. Perler, F. B. (1998) *Cell* **92**, 1–4
4. Xu, M.-Q., and Perler, F. B. (1998) *EMBO J.* **15**, 5146–5153
5. Perler, F. B. (1999) *Nucleic Acids Res.* **27**, in press
6. Wu, H., Hu, Z., and Liu, X. Q. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9226–9231
7. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Huang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadasoara, R., Vicaire, R., Wang, Y., Wierzbozki, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Church, G. M., Daniels, C. J., Mao, J.-I., Rice, P., Nolling, J., and Reeve, J. N. (1997) *J. Bacteriol.* **179**, 7135–7155
8. Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M.-Q. (1998) *J. Biol. Chem.* **273**, 10567–10577
9. Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Lence, V. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H., and Xu, M. Q. (1997) *Gene* **92**, 271–281
10. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) *Science* **266**, 776–779
11. Evans, J. T. C., Benner, J., and Xu, M.-Q. (1998) *Protein Sci.* **7**, 2256–2264
12. Muir, T. W., Sondhi, D., and Cole, P. A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6705–6710
13. Gimble, F. S. (1998) *Chem. Biol.* **5**, R251-R256
14. Holford, M., and Muir, T. W. (1998) *Structure* **6**, 945–949
15. Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W., and Hofnung, M. (1984) *J. Biol. Chem.* **259**, 10606–10613
16. Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S., and Tanaka, H. (1994) *J. Bacteriol.* **176**, 4465–4472
17. Maina, C. V., Riggs, P. D., Grandea, A. G., III, Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A., and Guan, C. D. (1988) *Gene (Amst.)* **74**, 365–373
18. Chong, S., Montello, G. E., Zhang, A., Cantor, E. J., Liao, W., Xu, M. Q., and Benner, J. (1998) *Nucleic Acids Res.* **26**, 6199–6215
19. Kluban, W., Sharma, S., Teleni, A., Jacobs, W. R. J., and Sachettini, J. C. (1998) *Nat. Struct. Biol.* **5**, 31–36
20. Teleni, A., Southworth, M., Alcaide, F., Dangela, S., Jacobs, J., William, R., and Perler, F. B. (1997) *J. Bacteriol.* **179**, 6378–6382
21. Wu, H., Xu, M.-Q., and Liu, X.-Q. (1998) *Biochim. Biophys. Acta* **1387**, 422–432
22. Pietrokovski, S. (1998) *Protein Sci.* **7**, 64–71
The in Vitro Ligation of Bacterially Expressed Proteins Using an Intein from Methanobacterium thermoautotrophicum

Thomas C. Evans, Jr., Jack Benner and Ming-Qun Xu

J. Biol. Chem. 1999, 274:3923-3926.
doi: 10.1074/jbc.274.7.3923

Access the most updated version of this article at http://www.jbc.org/content/274/7/3923

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 21 references, 8 of which can be accessed free at http://www.jbc.org/content/274/7/3923.full.html#ref-list-1