Protein splicing elements, or inteins, catalyze their own excision from flanking polypeptide sequences, or exteins, thereby leading to the formation of new proteins in which the exteins are linked directly by a peptide bond. A trans-splicing system, using separately purified N- and C-terminal intein fragments of about 100 amino acids each, fused to appropriate exteins, was recently derived from the *Mycobacterium tuberculosis* RecA intein (Mills, K. V., Lew, B. M., Jiang, S.-Q., and Paulus, H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 3543–3548). We have replaced the C-terminal intein fragment of this system with synthetic peptides comprising 35–50 of the C-terminal residues of the RecA intein. The N-terminal intein fragment and the synthetic peptide were reconstituted by renaturation from guanidinium chloride. In the absence of added reductants, a disulfide-linked dimer of the N-terminal fragment and the peptide accumulated and could be induced to splice by reduction of its disulfide bond. The intermediate and spliced products were identified by polyacrylamide gel electrophoresis, mass spectrometry, and derivatization with thiol-reactive biotin followed by Western blotting with a streptavidin-enzyme conjugate. This is the first example of protein splicing involving a synthetic intein fragment and opens the way for studying the active site structure and function of the intein by the use of different synthetic peptides, including ones with non-natural amino acids.

Protein splicing is a mechanism for the post-translational processing of proteins that involves the self-catalyzed excision of an intervening polypeptide, the intein, and the subsequent formation of a new protein by joining the flanking sequences, the exteins, by a peptide bond. It involves the catalysis of three mechanistically unrelated reactions at a single catalytic center, which resides entirely within the intein (for reviews, see Refs. 1 and 2). The intein can thus be viewed as an exceedingly complex enzyme, and the investigation of the catalytic mechanisms involved in protein splicing is of great interest.

With the aim of obtaining an in vitro protein splicing system whose structure and function can be examined by biochemical and biophysical methods, we are developing a minimal protein splicing element from the 440-residue RecA intein of *Mycobacterium tuberculosis* by eliminating the portions of the intein that are not essential for protein splicing. Most inteins are interrupted by homing endonuclease domains, which account for about one-half of the intein sequence but can be deleted without eliminating protein splicing ability (3–5). In addition, we found that the RecA intein can be split into two fragments that can complement each other so as to promote trans-splicing (5). This made possible the development of an in vitro trans-splicing system composed of 105-residue N- and C-terminal fragments of the *M. tuberculosis* RecA intein (6). An in vitro trans-splicing system based on the Pol-1 intein of the hyperthermophilic archeon, *Pyrococcus* sp. GB-D, was recently described (7). The results described in this paper further advance our approach by replacing the natural C-terminal intein fragment with 35–50-residue synthetic polypeptides. The resulting semisynthetic protein splicing element was able to catalyze protein splicing with high efficiency. This exciting development will facilitate the study of the structure and catalytic function of the C-terminal portion of the intein by replacing specific residues with other natural amino acids or with unnatural amino acids and structural probes.

**EXPERIMENTAL PROCEDURES**

**Plasmid Constructs and Protein Expression and Purification—**The construction of plasmid pmU2 s/sD6, which encodes MU<sub>N<sub>A</sub></sub> (an in-frame fusion of MBP to the 105 N-terminal amino acids of the *M. tuberculosis* RecA intein, followed by the C-terminal sequence Arg-Gly-Glu-Phe-COOH) was described earlier (5). MU<sub>N</sub> was expressed in *Escherichia coli* DH5<sub>a</sub> and purified as described previously (6).

**Peptide Synthesis—**Peptides with a C-terminal amide group ranging from 33 to 52 residues in length (see Fig. 1) were synthesized by N-(9-fluorenylmethoxycarbonyl) chemistry on an Applied Biosystems model 431 peptide synthesizer, using 4-methyl benzhydrylamine resin (Novabiochem, San Diego, CA), O-trityl-protected Thr, double coupling of all Thr and Arg residues, an acetic anhydride blocking step at each cycle, and N-methylpyrrolidone supplemented with dimethyl sulfoxide to 10%. The peptides were cleaved from the resin and deprotected with 6.25% (w/v) phenol in thioanisole:1:2-ethanedithiol:water:trifluoroacetic acid (3:2:1:20). The peptides were purified by HPLC (Rainin), using a preparative C8 column (Vydac) and monitoring the fractions by MALDI-TOF MS.

**MALDI-TOF MS—**Protein samples were dialyzed overnight against water and mixed on a mass spectrometry plate with an equal volume of 2,5-dihydroxybenzoic acid (1 mg/ml) in water:isopropyl alcohol:formic acid (3:2:1). An external standard of 1 mg/ml bovine serum albumin was also prepared, and the MU<sub>N</sub> starting material was used as an internal mass standard. MS was performed on a Voyager RP Biospectrometry.
Semisynthetic Protein Splicing Element

RESULTS

Ability of a Synthetic Peptide to Function in Protein Splicing—In the in vitro trans-splicing system developed earlier, 105-residue N- and C-terminal fragments of the M. tuberculosis RecA intein, fused to appropriate exteins, were mixed in 6 M urea or GdmCl and reconstituted by removing the denaturant by dialysis (6). The two intein fragments formed a heterodimer, which underwent efficient protein splicing under reducing conditions. Upon replacing the C-terminal intein fragment with a synthetic peptide corresponding to the 50 C-terminal intein residues, linked to Cys-Ala as the C-extein (Fig. 1), an analogous set of reactions was observed, as shown in Fig. 2 (lane 5), about 55% of MU Na and consistent in molecular mass with the putative spliced product, i.e. M linked to Cys-Ala (Fig. 2, lane 4). The overall splicing reaction proceeded somewhat more efficiently (60% yield based on MU Na) when TCEP was added at the beginning of the reconstitution procedure (Fig. 2, lane 3). Neither the 61-kDa intermediate nor the 43-kDa putative spliced product was observed when the peptide or MU Na was omitted from the reaction (Fig. 2, lanes 1 and 2).

Identification of Intermediates and Products—The putative heterodimeric intermediate and splicing product were further characterized by MS. Samples of the starting materials, the products of renaturation in the absence of TCEP, and the splicing reaction after addition of TCEP were analyzed by MALDI-TOF MS. The major molecular species that was observed in the starting mixture corresponded to MU Na (m/z = 55,064) and the 52-residue peptide (m/z = 6,051) (Fig. 3B). Upon renaturation under non-reducing conditions, an additional component was found with m/z of 61,100, in close agreement with that expected for the disulfide-linked heterodimer of MU Na and the 52-mer (m/z = 61,113) (Fig. 3C). Upon addition of TCEP, the m/z = 61,100 ionic species disappeared, and three new major ionic species appeared with m/z = 43,200, 12,000, and 5,900, consistent with the predicted m/z values of 43,293, 11,964, and 5,875 for the splicing product (M-Cys-Ala), the intein fragment (U Na), and the 50-residue peptide fragment, respectively (Fig. 3D).

The expected product of protein splicing, M-Cys-Ala, differs by a mass of only 193 mass units, corresponding to the dipeptide Cys-Ala, from M itself, which could have been produced from cleavage at the upstream splice junction (5). We therefore used an independent chemical assay for the identification of the putative protein splicing product. Because protein splicing leads to transfer of Cys-Ala to the C terminus of M, which itself contains no Cys residues, it should be possible to distinguish between M and the splicing product, M-Cys-Ala, by a method that specifically detects thiols. The products of the splicing reaction were treated with the thiol-reactive biotin-maleimide derivative, BMCC, which should specifically label all proteins containing Cys residues, including MU Na and the splicing product, M-Cys-Ala, but not free M. After SDS-PAGE and blotting onto nitrocellulose membranes, biotinylated proteins were detected using a streptavidin-alkaline phosphatase conjugate. In the complete splicing mixtures, a 43-kDa protein was the major biotin-labeled species (Fig. 4, lanes 1 and 3), whereas only MU Na was labeled in a mixture without the 52-residue peptide (Fig. 4, lane 5) and neither labeled component was observed when MU Na was omitted (Fig. 4, lane 7). No signal was observed with samples that had not been subjected to
biotinylation (Fig. 4, lanes 2, 4, 6, and 8). As another control, molecular mass markers including MBP were labeled with BMCC and developed on a Western blot. All marker proteins with free thiols were labeled, whereas MBP was not (data not shown). The observation that the 43-kDa protein produced in the complete system could be biotinylated with BMCC identified it as the spliced protein, M-Cys-Ala, rather than a cleavage product such as M.

Characteristics of the Peptide-dependent Splicing Reaction—Investigation of the role of prior denaturation on the reconstitution and splicing reactions showed that significant amounts of MU_NA peptide heterodimer were formed when MU_NA and the peptide were mixed in the absence of GdmCl, together with some higher aggregates (Fig. 5, lane 3), but that subsequent reduction yielded little spliced product (Fig. 5, lane 4) compared with reaction mixtures in which MU_NA and the peptide were reconstituted under denaturing conditions (Fig. 5, lanes 1 and 2). It is interesting that the heterodimer formed under non-denaturing conditions (Fig. 5, lane 3) failed to undergo efficient splicing, suggesting that productive interaction of the intein fragments to form a functional protein splicing active center requires prior unfolding of the polypeptide chains.

The experiments described in Figs. 2–5 were carried out with a nearly 10-fold molar excess of the 52-residue peptide. When the ratio of peptide to MU_NA was varied at a constant concentration (42 μM) of MU_NA, maximum conversion to spliced product (55%) was observed with an equimolar amount of peptide, suggesting a stoichiometric interaction of the two intein components (Fig. 6). The extent of conversion of MU_NA to spliced product roughly paralleled the extent of conversion to disulfide-linked heterodimer when this was measured separately (see Figs. 2, lanes 4 and 5 and Fig. 5, lanes 1 and 2). The extent of conversion of MU_NA to spliced product varied from 55 to 90% (for example, compare Figs. 2 and 4).

Effect of Peptide Length on Protein Splicing—Peptides comprising fewer than the 50 C-terminal amino acids of the protein splicing element were also examined for their ability to function in protein splicing. Each peptide was present in a 3-fold molar excess with respect to MU_NA. The results summarized in Fig. 7 show that a peptide corresponding to the 35 C-terminal amino acids of the intein was fully able to substitute for the 52-mer, whereas a peptide corresponding to the 31 C-terminal amino acids was inactive.

DISCUSSION

The results described in this paper demonstrate that a semisynthetic protein splicing element can effectively catalyze the complex series of reactions that lead to protein splicing. Our experimental system consisted of two fragments of an intein linked to appropriate exteins, which could be non-covalently reconstituted to form a functional protein splicing element; one intein fragment was a natural 105-residue protein segment and the other a synthetic 50-residue polypeptide. There have been other examples of active semisynthetic enzymes that can be reconstituted by the association of a natural and a synthetic fragment, the first being the reconstitution of ribonuclease S from S-peptide (residues 1–20) and S-protein (residues 21–124), which are produced by the cleavage of ribonuclease A with subtilisin (9). Replacement of the S-peptide with synthetic analogs yields functional semisynthetic ribonuclease derivatives (e.g., Ref. 10). Protein splicing elements should lend themselves especially well to reconstitution as semisynthetic enzymes because the protein splicing active center is composed of polypeptide sequences that correspond to the extreme ends of
The intein. Interspersed between these protein splicing sequences is often an extensive, functionally unrelated homing endonuclease domain (11), which imposes a spatial and temporal gap between their synthesis and assembly into a single functional domain. Indeed, natural N- and C-terminal fragments of the *M. tuberculosis* RecA intein, separately expressed and purified, were found to reconstitute and undergo protein splicing with high efficiency (6).

Our observation that synthetic polypeptides corresponding to between 35 and 50 of the C-terminal amino acids of the intein could effectively promote protein splicing offers an excellent opportunity for probing the structure and function of the protein splicing active center by substituting other amino acids or unnatural amino acid analogs at specific positions in the peptide. An especially attractive feature of our experimental system is that we can measure the reconstitution reaction separately from protein splicing by using mildly oxidizing conditions under which a refolded disulfide-linked heterodimer accumulates, which can subsequently be made to undergo quantitative conversion to the spliced products by reduction with TCEP. One can, therefore, study the effect of amino acid substitutions on the reconstitution reaction *per se*, i.e. the formation of a disulfide-linked complex, or on the protein splicing reaction *per se*, which occurs upon reduction of the disulfide-linked heterodimer. In addition, because the disulfide-linked heterodimer can be isolated as a stable protein, the structure of the protein splicing active center and its perturbation by amino acid substitution can be studied by various biophysical methods. The unusual nature of the protein splicing element as an enzyme should make such future investigations especially exciting.

One question that can be addressed immediately concerns the minimum size of the downstream intein fragment that is required for protein splicing. In a deletion analysis, Derbyshire *et al.* (3) found that protein splicing occurs *in vivo* when all but the last 35 C-terminal amino acids of the *M. tuberculosis* RecA intein are deleted but not after deletion of all but the last 31 residues. The 50-residue sequence used in most of our experiments is larger than the minimum size of the C-terminal intein fragment required for protein splicing. However, we could reconstitute a functional semisynthetic protein splicing element with a synthetic peptide corresponding to the 35 C-terminal residues of the *M. tuberculosis* RecA intein but not with one corresponding to the 31 C-terminal residues (Fig. 7). By synthesizing polypeptides of intermediate size, we should be able to define precisely the minimal length required for protein splicing.

**Acknowledgments**—We thank Kaori Shingledecker for help and advice, Shu-qin Jiang for preparing the plasmid, Anna Wong for synthesizing the peptides, and Paul Leavis and Betty Gowell for use of their preparative HPLC.

**REFERENCES**

1. Shao, Y., and Kent, S. B. H. (1997) *Chem. Biol.* 4, 187–194
2. Perler, F., Xu, M. Q., and Paulus, H. (1997) *Curr. Opin. Chem. Biol.* 1, 292–299
3. Derbyshire, V., Wood, D. W., Wu, W., Dansereau, J. T., Dalgaard, L. Z., and Belfort, M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 11466–11471
4. Chong, S., and Xu, M. Q. (1997) *J. Biol. Chem.* 272, 15587–15590
5. Shingledecker, K., Jiang, S. Q., and Paulus, H. (1998) *Gene (Amst.)* 207, 187–195
6. Mills, K. V., Lew, B. M., Jiang, S. Q., and Paulus, H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 3453–3458
7. Southworth, M. W., Adam, E., Panne, D., Byer, R., Kautz, R., and Perler, F. B. (1998) *EMBO J.* 17, 915–926
8. Laemmli, U. K. (1970) *Nature* 227, 680–685
9. Richards, F. M., and Vithayathil, P. J. (1959) *J. Biol. Chem.* 234, 1459–1464
10. Chaiken, I. M., Freedman, M. H., Lyerla, J. R., and Cohen, J. S. (1973) *J. Biol. Chem.* 248, 884–891
11. Duax, W., Gimbie, F. S., and Quiocho, F. A. (1997) *Cell* 79, 555–564
12. Reith, M. E., and Munholland, J. (1995) *Plant Mol. Biol. Rep.* 13, 333–335
13. Perler, F. B., Olsen, G. J., and Adam, E. (1997) *Nucleic Acids Res.* 25, 1087–1093