Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the Endonuclease Motifs*

(Received for publication, March 12, 1997)

Shaorong Chong and Ming-Qun Xu‡

From New England Biolabs, Inc., Beverly, Massachusetts 01915

The protein splicing element (intein) of the vacuolar ATPase subunit (VMA) of *Saccharomyces cerevisiae* catalyzes both protein splicing and site-specific DNA cleavage. It has been demonstrated that the conserved splice junction residues are directly involved in protein splicing and the central dodecapeptide motifs are required for DNA cleavage. To examine whether the splicing activity of the intein can be structurally separated from the endonuclease motifs, we made large in-frame deletions at the central region of the intein. We demonstrate for the first time that protein splicing can proceed efficiently after the removal of the central region of the intein including the endonuclease motifs. Our results suggest that the N- and C-terminal regions of the *Sce* VMA intein may form a separate domain that is not only catalytically sufficient for protein splicing but also structurally independent from the endonuclease domain.

Protein splicing is a post-translational processing event in which an internal segment, the intein, from a protein precursor catalyzes its own excision and concomitantly ligates the flanking regions, the exteins, to form a mature protein (1). It has been shown that inteins plus the first residue of the C-terminal flanking region (C-extein)3 contain sufficient structural and catalytic elements to direct splicing in the context of foreign proteins (2–6). In many cases, inteins belong to a family of site-specific endonucleases that cleave DNA in alleles lacking the inteins at a location called the homing site (7, 8). Sequence analyses of inteins from diverse organisms revealed seven conserved motifs (motifs A–G) (9). Motifs A and G contain a set of highly conserved residues at the two splice junctions. In vitro studies of protein splicing of the inteins from the thermostable DNA polymerase of *Pyrococcus* sp. GB-D and the 69-kDa vacuolar ATPase subunit of *Saccharomyces cerevisiae* (Sce VMA intein) have shown that these conserved splice junction residues play defined roles in the protein splicing pathway (4, 6, 10–13). Motifs C and E in the central region are referred to as the dodecapeptide motifs and are a characteristic feature of homing endonucleases (14, 15). However, it has been shown that protein splicing is independent of endonuclease function because an archaeal intein with a mutation that abolishes the endonuclease activity can still splice efficiently (16). The *Sce* VMA intein contains all seven conserved motifs (see Fig. 1A) (9). The *Sce* VMA intein functions as an endonuclease that cleaves the yeast genome at a single location and initiates a gene conversion process that results in the transfer of the intein gene to other yeast strains (8). The dodecapeptide motifs C and E are directly involved in the DNA recognition and cleavage reaction (17). The catalytic properties of the *Sce* VMA intein generated from the natural protein splicing process are indistinguishable from those of the recombinant form, indicating that the endonuclease function of the intein is independent of the protein splicing process (18). Because protein splicing of the *Sce* VMA intein involves the splice junction residues (in motifs A and G), which are separated from the endonuclease motifs (C and E) in the primary sequence by more than 100 amino acids (see Fig. 1A), questions have been raised concerning whether all 454 *Sce* VMA intein residues are required for protein splicing and whether the N- and C-terminal regions of the intein may contain sufficient structural and catalytic elements to catalyze protein splicing.

In this paper, we report the construction and characterization of large in-frame deletions within the *Sce* VMA intein that remove the central region of the intein including the dodecapeptide motifs. The deletion mutants were studied in a chimeric three-component fusion system in which the *Sce* VMA intein (Y) or its deletion mutants (ΔY) were fused in-frame between the *Escherichia coli* maltose-binding protein as the N-extein (M) and the *Bacillus circulans* chitin-binding domain (B) as the C-extein. The resulting fusion constructs (pMYB for the wild-type full-length intein and pΔMYB for the intein deletion mutants) were expressed in *E. coli*, and protein splicing was examined both in the crude cell extract and in the amylose-purified proteins. We demonstrate that protein splicing proceeds efficiently in the chimeric protein fusion context when 184 residues in the central region of the intein including the endonuclease motifs were replaced with flexible peptide linkers.

**EXPERIMENTAL PROCEDURES**

The procedures for cell culture, protein expression, and purification were the same as those described previously (6) except that the *E. coli* strain ER2526 (λ F proA B lac G ΔlacZM15 zff minT10 (KanR) (lacU2 supE44 relA1Δ rpoB1 Δlack15 endA1 spoT11 thi-1 ΔmcrC-mrr114::IS10, Elisabeth Raleigh, New England Biolabs) was used. The crude cell extracts and amylose-purified proteins were analyzed by SDS-PAGE, followed by Coomassie Blue staining and Western blot analysis. SDS-PAGE was performed in 12% Tris-Glycine gels (Novex, San Diego, CA). For Western blot analyses, the SDS-PAGE gels were blotted onto nitrocellulose membranes and analyzed by probing with polyclonal antibodies against the maltose-binding protein (New England Biolabs) or the *Sce* VMA intein (gift of Dr. F. S. Gimble) as described by Perler et al. (7). All enzymes are from New England Biolabs.

To construct pMYB to demonstrate protein splicing of the wild-type intein, pMB129 (19) was digested with *Bam*H1 and *Aeg*I and ligated with complementary oligonucleotides 5′-GATCCCAGGTTTGGTCGCTGCTC-TACGCAAACTG-3′ and 5′-CGGCTCAGGCCCACACATTGG-TTCA-CACAACCTTGC-GG-3′ to yield pMYB, in which the Asn-454 → Ala mutation in the C-terminal splice junction of pMB129 was reverted to the wild-type asparagine residue.

‡To whom correspondence should be addressed: New England Biolabs, Inc., 32 Tozer Rd., Beverly, MA 01915. Tel.: 508-927-5054 (Ext. 372); Fax: 508-921-1350.

© 1997 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 272, No. 25, Issue of June 20, pp. 15587–15590, 1997

This paper is available on line at http://www.jbc.org

15587
Deletions were made by the polymerase chain reaction (PCR) using pLitYP containing the intein sequence (6) as a template. Both the mutations scheme and the fusion constructs are shown in Fig. 1. The primer sequences are: L114, 5'-GGTCTGTGAGCCAGGCTTCTTTTGGCCACATCCAAATGCCACCCATAGGCATGGTATG-3' (for primer 1) and 5'-GGTCTGTGAGCCAGGCTTCTTTTGGCCACATCAAGCTGAACTTG-3' (for primer 2). Amplification was carried out for 20 cycles using a Perkin-Elmer thermal cycler at 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 4 min (Fig. 1B, step (1)). The product was digested with NheI and then self-ligated by T4 DNA ligase to form a circular plasmid pLitYP, which was subsequently amplified by transforming into E. coli ER2271 (Elisha Benjamin, New England Biolabs) (Fig. 1B, step (2)). The XhoI-BamHI fragments from pLitYP that contained the deletion mutation were ligated with XhoI and BamHI to replace the wild-type sequence, yielding pMYB (Fig. 1B, step (3)). Each of the pMYB vectors contained a specific intein deletion mutation (Fig. 1A). pMYB1 or pMYB2 was digested with NheI and AatII and ligated with the complementary oligomers 5'-CTAGGGTGGTACCTTGAAACCGGCTGTAACCCGTGGACACGCCTG-3' and 5'-CTAGGGTGGTACCTTGAAACCGGCTGTAACCCGTGGACACGCCTG-3' to yield pMYB1 (NG) or pMYB2/NG, in which a peptide linker sequence encoding Ala-Ser-Gly-Asp-Val (ASDV linker) was inserted into the intein deletion site (Fig. 1B, step (4)). pMYB2 was also digested with NheI and AatII and ligated with the complementary oligomers 5'-CTAGGGTGGTACCTTGAAACCGGCTGTAACCCGTGGACACGCCTG-3' and 5'-CTAGGGTGGTACCTTGAAACCGGCTGTAACCCGTGGACACGCCTG-3' to yield pMYB1 (NG) in which a peptide linker sequence encoding Ser-Gly-Ser-Gly-Ser (SG linker) was inserted into the intein deletion site (Fig. 1B, step (4)). All intein mutations were confirmed by DNA sequencing (New England Biolabs).

To test the ability of the splicing protein MB fusion proteins to bind chitin, 1 ml of amylase purified MB (0.5 mg/ml) was mixed with 0.5 ml of chitin resin (New England Biolabs). After 10 min of incubation at 4 °C, the chitin resin was pelleted by centrifugation and washed three times with column buffer containing 20 mM HEPES (pH 8.0), 0.5 M NaCl. The bound protein was eluted by 2% SDS and analyzed by SDS-PAGE as above.

**RESULTS**

Large in-frame deletions within the Sce VMA intein were created by a PCR method using two primers (primers 1 and 2, Fig. 1B) annealing at the deletion sites and pLitYP containing the Sce VMA intein sequence (6) as the template (Fig. 1B, step (1)). Both primers contained an NheI site that allowed self-ligation of the PCR products to generate plasmid pLitYP in which the intein sequence between the two primers was deleted (Fig. 1B, step (2)). This deletion intein sequence was then transferred to a pMYB fusion construct replacing the full-length intein sequence to yield pMYB (Fig. 1B, step (3)). Because primer 2 also contained a second restriction site AatII, a linker sequence encoding a short flexible peptide of multiple Asn and Gly residues (NG linker) or Ser and Gly residues (SG linker) was inserted between the NheI and AatII sites, resulting in pMYB (NG) or pMYB2/NG (Fig. 1B, step (4)). All fusion constructs and the positions of deletion are illustrated in Fig. 1A for details, see “Experimental Procedures”.

Protein splicing of the wild-type Sce VMA intein in the MB fusion generates the ligated exteins, MB, and the excised intein, Y. Due to the presence of non-native sequences between the intein (Y) and the exteins (M and B) in the MB fusion protein (6, 19), the ligated exteins, MB, are expected to have a molecular mass of about 51 kDa (approximately 4 kDa larger than the sum of the molecular masses of native MBP (M, 42 kDa) and CBD (B, 5 kDa)). The major component in the amylose-purified protein (Fig. 2A, lane 4) is a mixture of MB (51 kDa) and Y (50 kDa), they were not clearly separated on the SDS-PAGE of the crude cell extract (Fig. 2A, lane 2). The identifications of MB and Y were further verified by Western blot analysis using antibodies specific for the maltooligase-binding protein (anti-MBP) and the Sce VMA intein (anti-Sce) and by the ability of MB to bind chitin (detection of Y by anti-Sce in the crude extract is shown in Fig. 2A, lane 5).
may be the cleavage products of the MYB fusion proteins at a single splice junction (i.e. MY and M). The above data indicate that the wild-type Sce VMA intein does splice efficiently in the MYB fusion.

A deletion between residues 204 and 387 of the Sce VMA intein removed the dodecapeptide motifs (C and E) and motif D to yield pAMYB1 (Fig. 1B). Due to the incorporation of restriction sites in primer 1 and 2, six residues, Ala-Ser-Gly-Ala-Pro-Val, were inserted between residues 204 and 387 of the deletion intein. Expression of pAMYB1 in E. coli resulted in completely unspliced precursors (ΔMYB1) as shown in SDS-PAGE of the crude cell extract and amylose-purified proteins (Fig. 2B, left panel, lanes 1 and 2, respectively). No splicing products were observed when the purified ΔMYB1 precursors were incubated under splicing conditions (i.e. 20 mM HEPES (pH 8.0), 0.5 M NaCl) at 4 °C (Fig. 2B, left panel, lane 3) or 23 °C (data not shown). Similar results were obtained when deletions were made between residues 114 and 387 (ΔAMYB2) (data not shown).

A flexible 19-residue peptide linker containing Asn-Gly repeats (NG linker) or a 14-residue peptide linker containing Ser-Gly repeats (SG linker) was inserted into the deletion site of pAMYB1 to yield pAMYB1(NG) or pAMYB1(SG). After transformation of the plasmid into E. coli, the expressed proteins were purified on an amylose column. As shown in SDS-PAGE, the major component of the purified proteins was a 51-kDa protein, the size expected for the ligated exteins, MB (Fig. 2B, middle and right panel, lanes 2), and the excised intein mutant ΔY was detected in the crude extract with the expected molecular mass of ~33 kDa (Fig. 2B, middle and right panel, lanes 1), indicating that efficient splicing had occurred in vivo. The identities of MB and ΔY were further verified by Western blot analysis. The expected MB reacted with anti-MBP but not anti-Sce (data not shown), and it bound chitin as indicated by 2% SDS elution from the chitin resin (Fig. 2B, middle panel, lane 3). The anti-Sce antibodies reacted specifically with the excised intein mutant ΔY in the crude cell extract (Fig. 2B, middle panel, lane 4, and right panel, lane 3). Three minor components with high molecular masses were detected in amylose-purified proteins, the 2% SDS elutes, and they reacted with both anti-MBP and anti-Sce VMA intein sera (Fig. 2B, middle panel, lanes 3 and 4, and right panel, lane 3), suggesting that they were probably incomplete splicing products, i.e. the precursor (84 kDa), branched intermediate, and C-terminal splice junction cleavage product May (77 kDa). Other minor bands in the amylose-purified proteins, corresponding to the size of ΔY (33 kDa) and MBP (42 kDa), were also co-purified (Fig. 2B, middle and right panel, lane 2), indicating that some splicing and single splice junction cleavage of the precursor molecules occurred during purification and storage.

The NG linker was also inserted into the deletion site of pAMYB2 to yield pAMYB2(NG). Only a 74-kDa protein, corresponding to the unspliced precursor ΔMYB2(NG), was purified from the amylose affinity column (Fig. 2C, lane 2). Incubation of the purified ΔMYB2(NG) under splicing conditions (i.e. 20 mM HEPES (pH 8.0), 0.5 M NaCl) at 4 or 23 °C showed no splicing (data not shown).

**DISCUSSION**

For efficient splicing to occur, the protein precursor has to fold properly to bring the two splice junctions in close proximity and precisely align all the reacting groups. Previous studies have not determined whether the entire intein sequence including the functionally unrelated endonuclease motifs is required for the proper folding of the precursor that leads to efficient splicing. It has been shown that a small deletion in an
intein from the Vent DNA polymerase of Thermococcus litoralis resulted in an unspliced precursor (16), and a large deletion in the central region of an intein from Mycobacterium tuberculosis recA protein blocked splicing (2). In the case of the Sce VMA intein, although a seven-residue deletion in the middle of the Sce VMA intein (between motifs C and D) did not affect splicing, large in-frame deletions have been shown to block splicing (3). In the present study, large deletions in the intein central region also blocked splicing (pΔMYB1 and pΔMYB2). However, efficient splicing was observed when flexible peptide linkers were inserted at the deletion sites (pΔMYB1(NG) and pΔMYB1(SG)) (Fig. 2B). It appears that deletions in the central regions of the Sce VMA intein disrupt the remaining intein structure for proper folding thereby blocking splicing, albeit sufficient catalytic and structural elements for splicing still remain. Introduction of a linker at the deletion site appears to give flexibility to the local protein structure, thereby allowing the remaining intein to achieve the proper conformation for efficient splicing.

Protein splicing of the Sce VMA intein directly involves three essential splice junction residues, Cys-1, Asn-454, and Cys-455 (6). Mutations of intein residues close to the splice junctions have been shown to disrupt protein splicing (3, 16, 19, 20). A close examination of the chemical mechanism of protein splicing suggests that there may be other catalytic residues that are involved in assisting the nucleophilic attack by the splice junction cysteines or asparagine. A conserved histidine residue in motif B was suggested to be involved in assisting in the N- or C-terminal splicing reactions (9). Similarly, the data from this study also suggest that all catalytic residues are located in the splice junction motifs (A and G) and the proximal motifs (B and F), whereas the central motifs (C, D, and E) are not essential. Both ΔMYB1 (NG) and ΔMYB1 (SG) retained motifs A, B, F, and G and allowed efficient splicing. A motif H, spanning residues 340–359 in the Sce VMA intein, has recently been identified (21), and it is located within the deleted region of ΔMYB1 and presumably not required for splicing. It appears that the motifs that are involved in splicing may not be required for the endonuclease activity of the intein. Mutations of the conserved residues in motifs A, B, F, and G have no effect on the endonuclease activity. It is yet to be determined if an intein without the splice junction motifs A, B, F, and G can retain the endonuclease function. When a larger deletion was made as in ΔMYB2 (NG) (Fig. 2C), protein splicing was blocked even though the same peptide linker was inserted at the deletion site (Fig. 2C). It is possible that the deletion in ΔMYB2 (NG) contains certain essential elements for splicing or it disrupts the proper alignment of the catalytic residues. The intein deletion mutant ΔY in ΔMYB1(NG) or ΔMYB1(SG) also exhibited some differences in splicing efficiency from the full-length intein possibly due to certain structural alterations in ΔY caused by the deletion. For instance, compared with the MYB fusion, ΔMYB1(NG) and ΔMYB1(SG) fusion proteins spliced less efficiently as indicated by accumulation of a small amount of the precursors in the amylose-purified proteins (Fig. 2B, middle and right panel, lane 2). It is possible that the position of the deletion or the choice of peptide linker was not optimal for the remaining intein structure to achieve the wild-type splicing efficiency. Further variation in the deletion sites and/or linker sequences may optimize the splicing efficiency.

Despite the differences between ΔY and the wild-type intein, the data from this study suggest that ΔY retains not only most of the wild-type splicing activity but also the overall wild-type structure and that deleting the central region of the Sce VMA intein including the endonuclease motifs may not affect the overall structure of the remaining intein. This raises the possibility that in the tertiary structure of the Sce VMA intein, the N- and C-terminal regions may closely interact to form an independent “splicing domain,” whereas the central region of the intein may form an endonuclease domain. However, whether the endonuclease function of the Sce VMA intein resides only in the central region of the intein, thereby forming a separate domain, remains to be determined. A 150-amino acid open reading frame region in the chloroplast DnaB helicase protein of the red alga Porphyra purpurea and a 198-amino acid open reading frame in the gyrase A protein of Mycobacterium xenopi have been recently identified as putative inteins by sequence alignment (9, 21). Both Ppu dnaB (Fig. 1A) and Mxe gyrA inteins contain motifs A, B, F, and G but completely lack the dodecapeptide motifs (C and E) and motif D. Although these inteins have not been demonstrated to splice in vivo or in vitro, it is possible that naturally occurring inteins with only a splicing domain may exist in some proteins and that splicing occurs without the endonuclease motifs.

In conclusion, we have provided strong evidence that the N- and C-terminal regions of the Sce VMA intein including motifs A, B, F, and G contain sufficient structural and catalytic elements for splicing, whereas the central region of the intein including the dodecapeptide motifs C and E and motif D are not essential for protein splicing. Our data further our understanding of the mechanism of protein splicing and help to elucidate the three-dimensional structure of inteins. It also represents an important step toward “engineering” a minimal protein splicing element. Although a rational design of such a minimal splicing element awaits solving of the intein crystal structure, our study provides a new approach to intein-based protein engineering and a variety of applications in molecular biology.

Acknowledgments—We thank D. Comb, C. Noren, G. A. Garcia, W. Jack, F. Perler, F. Mersha, M. Southworth, H. Paulus, T. Evans, Jr., D. Shub, and M. Scott for valuable discussions and reading of the manuscript.

Note Added in Proof—The crystal structure of the Sce VMA intein (Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) Cell 90, 555–564) shows that the intein is composed of two separate domains, which is consistent with our experimental data and two domain model.

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. Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J., and Sedgwick, S. G. (1994) Cell 71, 201–210.
3. Cooper, A. A., Chen, Y., Linderfer, M. A., and Stevens, T. H. (1993) EMBO J. 12, 2575–2583.
4. Xu, M.-Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Perler, F. B. (1993) Cell 73, 1371–1377.
5. Davis, E. O., Thangaraj, H. S., Brooks, P. S., and Colston, M. J. (1994) EMBO J. 13, 699–703.
6. Chong, S., Yang, S., Paulus, H., Jannasch, H. S., and Jannasch, H. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 5577–5581.
7. Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., Benner, J., Slabko, B. E., Gorned, D. O., Hemphard, S. K., Carlow, C. K. S., and Jannasch, H. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 5577–5581.
8. Gimble, F. S., and Thorner, J. (1992) Nature 357, 301–306.
9. Pietruszkowski, S. (1994) Protein Sci. 3, 2340–2350.
10. Shao, Y., Xu, M.-Q., and Paulus, H. (1995) Biochemistry 34, 10844–10850.
11. Shao, Y., Xu, M.-Q., and Paulus, H. (1996) Biochemistry 35, 3810–3815.
12. Xu, M.-Q., Comb, D. G., Paulus, H., Noren, C. J., Shao, Y., and Perler, F. B. (1994) EMBO J. 13, 5517–5522.
13. Xu, M.-Q., and Perler, F. B. (1996) EMBO J. 15, 5146–5153.
14. Shub, D. A., and Goodrich-Blair, H. (1992) Cell 71, 183–186.
15. Lymbouzis, A. M., and Belfort, M. (1990) Annu. Rev. Biochem. 60, 587–622.
16. Hodges, R. A., Perler, F. B., Noren, C. J., and Jack, W. E. (1992) J. Biol. Chem. 268, 7122–7128.
17. Lymbouzis, A. M., and Belfort, M. (1990) Annu. Rev. Biochem. 60, 587–622.
18. Hodges, R. A., Perler, F. B., Noren, C. J., and Jack, W. E. (1992) J. Biol. Chem. 268, 7122–7128.
19. Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, B. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H., and Xu, M.-Q. (1997) Gene (Amst.), in press.
20. Hirata, R., and Anraku, Y. (1992) Biochem. Biophys. Res. Commun. 188, 40–47.
21. Perler, F. B., Olsen, G. J., and Adam, E. (1997) Nucleic Acids Res. 25, 1087–1093.

2 Vence and Xu, unpublished results.