Protein splicing in the absence of an intein penultimate histidine.

Lixin Chen, Jack Benner and Francine B. Perler
New England BioLabs Inc., 32 Tozer Road, Beverly, MA 01915, USA

978-927-5054, ext.: 326
Fax: 978-921-1350
email: perler@neb.com

Running Title: Penultimate His
SUMMARY

Protein splicing is a self-catalytic process in which an intervening sequence, termed an intein, is excised from a protein precursor and the flanking polypeptides are religated. The conserved intein penultimate His facilitates this reaction by assisting in Asn cyclization, which results in C-terminal splice junction cleavage. However, many inteins do not have a penultimate His. Previous splicing studies with 2 such inteins yielded contradictory results. To resolve this issue, the splicing capacity of 2 more inteins without penultimate His residues was examined. Both the Methanococcus jannaschii phosphoenolpyruvate synthase and RNA polymerase subunit A' inteins spliced. Splicing of the phosphoenolpyruvate synthase intein improved when its penultimate Phe was changed to His, but splicing of the RNA polymerase subunit A' intein was inhibited when its penultimate Gly was changed to His. We propose that inteins lacking a penultimate His (i) arose by mutation from ancestors in which a penultimate His facilitated splicing, (ii) that loss of this His inhibited, but may not have blocked splicing and (iii) that selective pressure for efficient expression of the RNA polymerase yielded an intein which utilizes another residue to assist Asn cyclization, changing the intein active site so that a penultimate His now inhibits splicing.
INTRODUCTION

Protein splicing is a post-translational event analogous to RNA splicing that involves the removal of an internal protein fragment (intein) from a precursor protein and the joining of the two flanking sequences (exteins) to produce an active extein protein (1). As of January 2000, 100 putative inteins have been identified in all three phylogenetic domains (see the Intein Registry in InBase (2) at <http://www.neb.com/neb/inteins.html>). Inteins have 2 structural domains (Fig. 1A): a splicing domain which is composed of the N-terminal and C-terminal splicing regions, and a central region encoding a homing endonuclease (3,4) or a small linker. The structure of the intein splicing domain is conserved amongst inteins, forming a protein fold termed the HINT module (5-8). The homing endonuclease domain is dispensable. Residues mediating or assisting the protein splicing reaction have been identified by sequence comparison, mutation and structural analysis (6,8-19) (Fig. 1A).

The protein splicing reaction requires the intein splicing domain plus the first amino acid (aa1) of the C-extein. The self-catalytic protein splicing mechanism (reviewed in Noren et al., (20)) consists of four steps, each involving nucleophilic displacements (Fig. 2): (i) formation of a linear (thio)ester by an acyl rearrangement of the conserved Cys1 or Ser1 at the intein N-terminus, (ii) formation of a branched intermediate by transesterification when the thiol/hydroxyl group of the Ser, Thr or Cys at the beginning of the C-extein attacks the (thio)ester from Step 1, resulting in transfer of the N-extein to the side chain of the first C-extein residue, (iii) resolution of the branched intermediate by cyclization of the intein C-terminal Asn or Gln, resulting in cleavage of the peptide bond between the intein and the C-extein, and (iv) formation of a native peptide bond between the exteins by a spontaneous S-N (or O-N) acyl rearrangement. Dead-end side reactions are often observed when the intein is expressed in a heterologous extein. N-terminal splice junction cleavage products are formed by cleavage of the (thio)ester at either the N-terminal splice junction or at the branch point in the branched intermediate. C-terminal splice junction cleavage results when Asn or Gln cyclization precedes Steps 1 or 2. Although inteins are not true enzymes in the
sense that they do not act on multiple substrates, they use the same mechanisms as enzymes to facilitate splicing, including oxyanion holes at each splice junction (6,20).

This paper focuses on the role of the intein penultimate residue, which is thought to facilitate C-terminal splice junction cleavage and Asn cyclization when His is present at this position. Inteins must increase the rate of Asn or Gln cyclization, since Asn cyclization takes several days at 37°C, pH 7.4 in model peptides and cyclization of Gln is even slower due to entropic factors (21). Moreover, in both proteins and peptides, the preferred reaction is the attack by the main chain nitrogen on the side chain carbonyl leading to deamidation, as opposed to the protein splicing reaction in which the side chain amide nitrogen attacks the main chain carbonyl (21). Inteins can increase the rate of Asn or Gln cyclization and direct the reaction towards peptide bond cleavage by aligning reactive groups and rendering the peptide bond more labile by generating an electrophilic center at the carbonyl carbon. The increased electrophilicity could be accomplished by (i) hydrogen bonding to the carbonyl oxygen, thereby stabilizing the developing negative charge on the tetrahedral intermediate of the cyclization reaction and/or (ii) making the main chain amine a better leaving group by donating a proton. Inteins could also facilitate cyclization by increasing the nucleophilicity of the side chain amide group. The intein penultimate His is thought to assist Asn cyclization. Most inteins have a His in the penultimate position (2) and mutation of this His inhibits or blocks splicing (10,11,18). The crystal structures of the *Mycobacterium xenopi* gyrase subunit A (GyrA) intein (6) and the *Saccharomyces cerevisiae* ATPase (VMA) intein (8) in the absence of a C-extein reveal that the intein penultimate His is in hydrogen bonding distance to an Asn carboxylate oxygen. The penultimate His therefore facilitates Asn cyclization by making the Asn carbonyl carbon more electrophilic. Modeling of the *M. xenopi* GyrA intein suggests that the penultimate His may donate a proton to the main chain amine leaving group (6).

Seventeen inteins (10 families of intein alleles) have another residue at the penultimate position (2), including Gly, Ala and Phe which are unlikely to assist in Asn or Gln cyclization (Fig. 1B). Are these inteins capable of splicing? Previous studies of 2 such inteins yielded
contradictory results. The *Chlamydomonas eugametos* ClpP intein failed to splice in *Escherichia coli* unless its penultimate Gly was mutated to His (19), suggesting that inteins lacking a penultimate His are inactive. On the other hand, the DnaE intein of *Synechocystis sp.* PCC6803 spliced in *E. coli* (22), although more spliced product accumulated when its penultimate Ala was mutated to His (23). To further address this issue, the *Methanococcus jannaschii* (Mja) phosphoenolpyruvate synthase (PEP) and RNA polymerase subunit A' (Rpol A') inteins with Phe and Gly at the intein penultimate position, respectively, were tested for their ability to splice with their native penultimate residue and with a penultimate His.
EXPERIMENTAL PROCEDURES

Construction of MEIEP clones. The Pyrococcus sp. GB-D DNA polymerase intein in pMIP21 (24) was replaced with either the Mja Rpol A' intein or the Mja PEP intein plus varying lengths of native N- and C-extein sequences (E) to generate MEIEP fusions (Fig. 1C). DNA encoding the M. jannaschii inteins was amplified by polymerase chain reaction (PCR) from genomic DNA using Mja Rpol A' primers, 5'-ATG CTC GAG CAC CAA ACT TAA TTA TTG AGG AT and 5'-ATT GAG GCC TTC AGC GTT GTA GGG CGG GCA ATT or Mja PEP primers, 5'-ATG CTC GAG ACG ATG TTT GTT AAG GAT GAA AAA and 5'-ATT GAG GCC TCT AGA AAC GAT TGC CGC GTG GCA GTT. The Mja Rpol A' fragment included 200 native N-extein residues and 5 native C-extein residues. The Mja PEP intein fragment included 150 native N-extein residues and 7 native C-extein residues. The resultant PCR fragment was digested with XhoI and StuI, purified from agarose gels with a QIAEX II gel extraction kit (Qiagen Inc.) and ligated into gel-purified XhoI/StuI cut pMIP21 vector DNA. All mutants were constructed by substituting a restriction enzyme fragment in pMEIEP with a PCR fragment generated using the following primers containing the desired mutations: PEP intein F411H primers 5'-AAA TGG AAA CCA ATA AGG GT and 5'-GCG GCC CTA AAA CGA TTG CCG CGT GGC AGT TAT GTA CAA CTA TTGG; Rpol A' intein G451H primers 5'-GAA TTT GGT ATT GAA TTA AAG and 5'-ATT GAG GCC TTC AGC GTT GTA GGG CGG GCA ATT GTG TGT TAA AAA GCC; Rpol A' intein primers 5'-AAA GAG AAA TGC CTT AA GAA TGGA and 5'-CCT TCA GCG TTG TAG GGC GGG CAA TTX XXT GTT AAA AAG CC, in which XXX was TGC for G451A, AAA for G451F, TTT for G451K and CCA for G451W. A His$_6$ tag was incorporated at the end of paramyosin by insertion of a double-strand oligonucleotide into the SalI and PstI sites in the PEP intein clone and the SalI and HindIII sites in the Rpol A' intein clone.

The sequence of all PCR fragments was confirmed after sequencing both DNA strands by the NEB DNA sequencing core facility. All enzymes were obtained from New England BioLabs and used as described by the manufacturer.
MEIEP production and characterization. *E. coli* strain TB1 (New England BioLabs) cells containing pMEIEP plasmids were grown at 30°C to mid log phase and then further incubated in the presence or absence of 0.4 mM IPTG at 15°C overnight. Nickel column (Qiagen Inc.) purified proteins containing the C-terminal His tag were prepared as described by the manufacturer with the addition of 10% glycerol in all buffers. Induction of MEIEP precursors in the presence of IPTG resulted in increased proteolytic cleavage within the intein and little or no increase in precursor, spliced products or single splice junction cleavage products. Therefore, experiments were performed with purified samples from uninduced cultures.

MEIEP precursors, spliced products and cleavage products are distinguishable by relative mobility in Coomassie Blue stained SDS-PAGE and by immuno-reactivity in Western blots using anti-MBP and anti-paramyosin sera (24). In order to obtain clear separation of MEEP and IEP, samples were electrophoresed in 8% acrylamide SDS-PAGE and the 30 kDa EP product was run off the bottom of the gel. The same samples were also electrophoresed in 12% acrylamide SDS-PAGE and stained with Coomassie Blue for analysis of the EP product (data not shown). Coomassie Blue stained gels were digitized with a Microtek Scanmaker III and quantitated with NIH Image 1.51 software as described (25). The values for at least 3 independent experiments were averaged for each sample and the standard deviation of the means calculated. The increase or decrease in molar percent of spliced product (MEEP) or N-terminal splice junction cleavage product (IEP) was calculated as follows: percent increase = (mutant - wild type)/wild type x 100 and percent decrease = (wild type - mutant)/wild type x 100.

N-terminal sequencing by Edman Degradation. Purified Mja Rpol A' MEIEP samples were sequenced as described (24). Briefly, protein samples were subjected to electrophoresis on a 8% Tris-Glycine polyacrylamide gel (Novex) and transferred to a ProBlott PVDF membrane (ABI). The membrane was stained with Coomassie Blue R-250, bands corresponding to the branched intermediate (MEIEP*), the MEIEP precursor and the MEEP spliced product were excised, and
each subjected to sequential Edman degradation. The data was acquired and analyzed on an
Applied Biosystems 610A Data System.
RESULTS

Splicing of the Mja PEP and Mja Rpol A’ inteins in E. coli. The well studied MIP in vitro protein splicing system (18,24-26) was used to examine whether a penultimate His is required for splicing of the Mja Rpol A’ and PEP inteins. The MIP precursor consists of an intein (I) inserted in-frame between the E. coli Maltose Binding Protein (MBP or M) and the Δ Sal fragment of Dirofilaria immitis Paramyosin (P). In this study, native extein sequences (E) were included to improve splicing by better mimicking the native precursor active site, generating the MEIEP fusions (Fig. 1C). The Mja PEP intein has a penultimate Phe at position 411 and the Mja Rpol A’ intein has a penultimate Gly at position 451. Size, immuno-reactivity and, in some cases, N-terminal protein sequencing were used to identify precursor (MEIEP), spliced product (MEEP), branched intermediate (MEIEP*) and N-terminal splice junction (ME + IEP) or C-terminal splice junction (MEI + EP) cleavage products. Cleared cell lysates were chromatographed over nickel chelating columns, resulting in purification of proteins containing the C-terminal His tag and loss of ME and MEI. The Mja PEP and Rpol A’ MEIEP precursors spliced efficiently in E. coli (Fig. 3), with 60% spliced MEEP product observed with the Mja PEP intein and 68% with the Mja Rpol A’ intein (Table I). No C-terminal splice junction cleavage product (EP) was observed when samples were electrophoresed in 12% SDS-PAGE (data not shown). No increase in spliced product was observed if protein samples were incubated overnight in vitro at 16°C or 37°C at pH 6.0 to 8.5 (data not shown). It is not known why some proportion of precursors usually fail to splice in heterologous systems, although it has most often been attributed to misfolding or aggregation.

Protein splicing of Mja PEP and Mja Rpol A’ inteins in MEIEP after mutation of the penultimate intein residue to His. Mutation of the Mja PEP intein penultimate Phe and the Mja Rpol A’ intein penultimate Gly to His was performed to see if ‘reversion’ to this normally conserved residue would improve splicing. There was a 43% increase in spliced MEEP product obtained from the F411H mutant Mja PEP intein samples as compared to the wild type samples (Fig. 3 and Table I). However, replacing the Mja Rpol A’ intein penultimate Gly with His resulted in a 27% decrease in
spliced product, a 57% increase in IEP N-terminal splice junction cleavage product and an accumulation of a slowly migrating protein; amino acid sequencing of the latter protein indicated that it was the branched intermediate, since it had the predicted pair of amino acids present in each of 15 cycles of Edman degradation (Table II). One amino acid in each cycle corresponded to the MBP sequence, while the other residue in that cycle corresponded to the Mja Rpol A' intein sequence. The identity of the Mja Rpol A' precursor (MEIEP) and spliced product (MEEP) were also confirmed by protein sequencing (data not shown). The accumulation of branched intermediate and N-terminal splice junction cleavage products are indicative of inhibiting only the Asn cyclization step and not the acyl shift or transesterification steps (Fig. 2).

*Further substitutions of the Mja Rpol A' intein penultimate residue.* The Mja Rpol A' intein penultimate residue, G451, was replaced with Ala, Lys, Phe, or Trp in the MEIEP context (Fig. 4 and Table III). Ala, Lys and Phe substitutions resulted in a similar decrease in spliced product (~27%) as in the G451H mutation. G451H and G451F resulted in an increase in N-terminal splice junction cleavage products and branched intermediate, suggesting that His and Phe substitutions only inhibited Asn cyclization. The bulky penultimate Trp residue resulted in the largest reduction in spliced product (59%), but still allowed N-terminal splice junction cleavage and the accumulation of branched intermediate. No C-terminal splice junction cleavage product (EP) was observed in any sample (data not shown).
DISCUSSION

A His is present at the penultimate position in 83% of inteins (2). Mutagenesis and structural studies have demonstrated the importance of this residue in facilitating Asn cyclization and protein splicing (6,8,10-12,18). Contradictory results were obtained when the first two inteins naturally lacking a penultimate His were tested for the ability to splice. The *C. eugametos* ClpP intein with a penultimate Gly (19) failed to splice in *E. coli*, while the *Synechocystis sp.* PCC6803 DnaE intein with a penultimate Ala did splice (22). The present study examined 2 more inteins that lack a penultimate His. Both the *M. jannaschii* PEP and Rpol A' inteins with a Gly or Phe at the penultimate position yielded >60% spliced product in *E. coli* despite the fact that neither Gly nor Phe can chemically participate in Asn cyclization. In experiments to be published elsewhere, we have also demonstrated splicing of the *Pyrococcus sp.* GB-D KlbA intein, which has a penultimate Ser.² To date, half of the intein families that lack a penultimate His have been examined and all but 1 are capable of splicing. We therefore suggest that the failure of the *C. eugametos* ClpP intein to splice in *E. coli* (19) is the exception, rather than the rule and that most inteins that lack a penultimate His will probably be active. It is possible that the *C. eugametos* ClpP intein splices in its native organism, since there is precedent for failure of active inteins to splice in *E. coli* (27).

The splicing reaction requires that the two splice junctions are in close proximity. Inteins are thought to have a complex active site with two oxyanion holes: one at the C-terminal splice junction that facilitates Asn or Gln cyclization and includes the intein penultimate His, and a second oxyanion hole at the N-terminal splice junction that facilitates the first two nucleophilic displacements and includes residues in intein Block B (6,20). Several substitutions in the penultimate position of the Mja Rpol A' intein had little effect on reactions at the N-terminal splice junction, since N-terminal splice junction cleavage products were observed and the branched intermediate accumulated. However, all of these mutations inhibited Asn cyclization as evidenced by the failure to rapidly resolve the branched intermediate and the absence of C-terminal splice junction cleavage products. Although these data support the presence of separable oxyanion holes
at each splice junction, precursor and intermediate structures would help to determine whether they are distinct or overlapping.

Both the *C. eugametos* ClpP and Mja Rpol A' inteins have a penultimate Gly. Unlike the *C. eugametos* ClpP intein which is activated by a penultimate His (19), the Mja Rpol A' intein is inhibited by a penultimate His. A penultimate His in the Mja Rpol A' intein may cause steric inhibition at the intein active site or block access to the C-terminal splice junction by a new residue(s) that facilitates Asn cyclization. Four more mutations were made at G451 ranging from the second smallest amino acid, Ala, to the largest amino acid, Trp. Substituting Gly with Ala yielded a similar reduction in spliced product as His, Lys and Phe, indicating that either packing at the C-terminal splice junction is so tight that substituting Gly with the slightly larger Ala will cause the same steric effect as substituting with Phe or that the conformational flexibility that Gly provides is critical for aligning active site residues at the Mja Rpol A' intein C-terminal splice junction.

We suggest that all inteins evolved from ancestors that had a penultimate His which facilitated Asn or Gln cyclization. Mutation of the intein penultimate His might have yielded an intein that still spliced, although less efficiently (unless other compensatory mutations occurred simultaneously). Asn or Gln cyclization would then become rate limiting. Enough active extein might still be synthesized to permit survival of the host until further mutation increased splicing proficiency. In support of this hypothesis, splicing of the *C. eugametos* ClpP (19), the *Synechocystis sp.* PCC6803 DnaE (23) and the Mja PEP inteins improved when the intein penultimate residue was 'reverted' to His. These inteins may yet to have fully compensated for the absence of a penultimate His. They may have recently mutated or selective pressure may be low since they may produce sufficient amounts of spliced extein despite inefficient splicing. The more essential and highly expressed the gene product, the stronger the selection for mutations that would improve splicing. Although *dnaE* encodes an essential protein that is part of the replicative DNA polymerase, splicing of *Synechocystis sp.* PCC6803 DnaE does not have to be very efficient, since only a few molecules of replicative DNA polymerase are generally required per cell (≤20
molecules/cell in *E. coli* (28). However, Rpol A’ is not only an essential protein, but it is a highly expressed protein, comprising part of the archaeal RNA polymerase. Therefore, individuals that acquired mutations that improved splicing of the Mja Rpol A’ intein would rapidly become fixed in the population. In fact, the Mja Rpol A’ intein has changed so much that splicing is now inhibited when its penultimate residue is 'reverted' back to His. The differences in splicing capacity of inteins that naturally lack a penultimate His may thus reflect inteins at different stages of evolving towards rapid splicing after mutation of their penultimate His. The data also suggest that splicing of inteins that naturally lack a penultimate His may improve if the native penultimate residue is replaced by His.
ACKNOWLEDGMENT

We thank Gary Olsen and David Graham, University of Illinois, for kindly providing us with *M. jannaschii* genomic DNA. DNA sequencing was performed by the New England BioLabs Core Sequencing Facility and protein sequencing was performed by Jack Benner. We thank Don Comb for support and encouragement, and Maurice Southworth, Eric Adam, Karen Sandman, Tom Evans and Chris Noren for helpful discussions and reading of the manuscript.
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. (2000) *Nucleic Acids Res.* **28**, 344-345
3. Belfort, M., and Roberts, R. J. (1997) *Nucleic Acids Res.* **25**, 3379-3388
4. Jurica, M. S., and Stoddard, B. L. (1999) *Cell. Mol. Life Sci.* **55**, 1304-1326
5. Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) *Cell* **91**, 85-97
6. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., Jr., and Sacchettini, J. C. (1998) *Nat. Struct. Biol.* **5**, 31-36
7. Perler, F. B. (1998) *Cell* **92**, 1-4
8. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) *Cell* **89**, 555-564
9. Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M. Q. (1996) *J. Biol. Chem.* **271**, 22159-22168
10. Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M. Q. (1998) *J. Biol. Chem.* **273**, 10567-10577
11. Cooper, A. A., Chen, Y., Lindorfer, M. A., and Stevens, T. H. (1993) *EMBO J.* **12**, 2575-2583
12. Hirata, R., and Anraku, Y. (1992) *Biochem. Biophys. Res. Comm.* **188**, 40-47
13. Kawasaki, M., Nogami, S., Satow, Y., Ohya, Y., and Anraku, Y. (1997) *J. Biol. Chem.* **272**, 15668-15674
14. Perler, F. B., Olsen, G. J., and Adam, E. (1997) *Nucleic Acids Res.* **25**, 1087-1093
15. Pietrokovski, S. (1994) *Protein Sci.* **3**, 2340-2350
16. Pietrokovski, S. (1998) *Protein Sci.* **7**, 64-71
17. Dalgaard, J. Z., Moser, M. J., Hughney, R., and Mian, I. S. (1997) *J. Comput. Biol.* **4**, 193-214
18. Xu, M., and Perler, F. B. (1996) *EMBO J.* **15**, 5146-5153
19. Wang, S., and Liu, X. Q. (1997) J. Biol. Chem. 272, 11869-11873
20. Noren, C. J., Wang, J., and Perler, F. B. (2000) Angewandte Chemie Int. Ed. 39, 450-466
21. Stephenson, R. C., and Clarke, S. (1989) J. Biol. Chem. 264, 6164-6170
22. Wu, H., Hu, Z., and Liu, X. Q. (1998) Proc. Natl. Acad. Sci. USA 95, 9226-9231
23. Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C., and Benkovic, S. J. (1999) Proc. Natl. Acad. Sci. U S A 96, 13638-13643
24. Xu, M., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Perler, F. B. (1993) Cell 75, 1371-1377
25. Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W. R., Jr., and Perler, F. B. (1997) J. Bacteriol. 179, 6378-6382
26. Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q., and Perler, F. B. (1999) BioTechniques 27, 110-120
27. Davis, E. O., Thangaraj, J. S., Brooks, P. C., and Colston, M. J. (1994) EMBO J. 13, 699-703
28. Kornberg, A., and Baker, T. A. (1992) DNA Replication, Second Edition Ed., W. H. Freeman and Co., New York
29. Pietrokovski, S. (1998). Curr. Biol. 8, R634-635.
Footnotes:

1 Abbreviations: aa, amino acid; Mj a, Methanococcus jannaschii; PEP, phosphoenolpyruvate synthase; Rpol A', RNA polymerase subunit A'; VMA, vacuolar ATPase; GyrA, gyrase subunit A; M and MBP, maltose binding protein; P, Dirofilaria immitis paramyosin Δ Sal fragment; I, Rpol A' or PEP intein; E, native extein; MEIEP, a chimeric precursor consisting of MBP, a partial native N-extein, an intein, a partial native C-extein and D. immitis paramyosin Δ Sal fragment; MEIEP*, the branched intermediate of the MEIEP precursor; MEEP, the spliced exteins from the MEIEP precursor; IEP, an N-terminal splice junction cleavage product of MEIEP; PCR, polymerase chain reaction.

2 M. W. Southworth and F. B. Perler, unpublished data
Figure Legends

Fig. 1A. Organization of a typical protein splicing precursor. A precursor is depicted with an intein composed of a homing endonuclease domain (black box) separating the N-terminal and C-terminal splicing regions (white boxes) which form the intein splicing domain. Mini inteins have a short linker instead of an endonuclease domain. Intein motifs are shown above the precursor and conserved residues in selected Blocks are shown below the precursor. Block A contains the conserved Ser or Cys at the N-terminus of the intein, although Ala is present at this position in 3 intein families. Block G contains the conserved dipeptide His-Asn at the intein C-terminus, and Ser, Thr or Cys at the beginning of the C-extein. Three inteins have Gln instead of Asn in Block G (2, 29).

Fig. 1B. Intein polymorphism at the penultimate residue. As of January 2000, 10 families of intein alleles lack a penultimate His. The number in parenthesis is the total number of residues in each intein. Symbols: Ceu, C. eugametos; CIV, Chilo iridescent virus; Aae, Aquifex aeolicus; Spb, B.subtilis SP beta phage; Ssp, Synechocystis sp. PCC6803; Pab, Pyrococcus abyssi; Pho, Pyrococcus horikoshii OT3; Pfu, Pyrococcus furiosus; RIR1, Ribonucleoside-diphosphate reductase, alpha subunit; KlbA, kilB operon orfA; LHR, Large helicase related protein; Lon, ATP-dependent protease LA; Moaa, Molybdenum cofactor biosynthesis homolog.

Fig. 1C. Schematic representation of the Rpol A' and PEP MEIEP precursors. Splice junction residues, the position of the intein penultimate amino acid (arrow) and the number of extein residues are indicated. Native extein sequences are represented by E and H represents a 6 aa His tag.

Fig. 2 The self-catalytic protein splicing reaction. The protein splicing mechanism is depicted with X representing either the oxygen or the sulfur present in the side chain of Ser, Thr or Cys. In some
inteins, Asn is replaced by Gln, which can similarly cyclize. All tetrahedral intermediates, assisting
groups and proton transfer steps are omitted for clarity.

**Fig. 3.** *Splicing of Mja Rpol A' and PEP inteins in MEIEP.* Mja Rpol A’ and PEP MEIEP
precursors containing the native intein penultimate residue or mutants in which the penultimate
position was ‘reverted’ to the consensus His residue (G451H or F411H, respectively) were
examined for *in vivo* splicing. Nickel column purified Mja Rpol A' or Mja PEP MEIEP samples
were electrophoresed in SDS-PAGE and the gels were stained with Coomassie Blue. Mja Rpol A'
precursor (MEIEP, 145 kDa), branched intermediate (MEIEP*), spliced product (MEEP, 105
kDa), and an N-terminal splice junction cleavage product (IEP, 80 kDa) were observed, as were
Mja PEP precursor (MEIEP, 135 kDa), spliced product (MEEP, 95 kDa), and an N-terminal splice
junction cleavage product (IEP, 75 kDa).

**Fig. 4.** *The effect of the intein penultimate residue on protein splicing of the Mja Rpol A' intein in
the MEIEP fusion protein.* The Mja Rpol A' intein penultimate residue (G451) was mutated in
MEIEP to Ala (G451A), His (G451H), Lys (G451K), Phe (G451F) or Trp (G451W) and samples
were analyzed as in Figure 3. Although all samples yielded spliced product, the percentage of
spliced product and branched intermediate varied with each mutation.
Table I

*Splicing and Cleavage Activities of the Mja Rpol A’ and PEP inteins in MEIEP*

| Protein | Rpol A’ | | PEP | |
|---------|---------|-----------------|---------|
|         | G451    | G451H           | F411    | F411H |
| MEIEP*  | 0\(^a\)  | 6 ± 2           | 0       | 0      |
| MEIEP   | 11 ± 1   | 10 ± 3          | 31 ± 12 | 7 ± 3  |
| MEEP    | 68 ± 4   | 50 ± 10         | 60 ± 6  | 86 ± 7 |
| IEP     | 21 ± 4   | 33 ± 6          | 9 ± 6   | 8 ± 6  |
| EP      | 0        | 0               | 0       | 0      |

\(^a\) Values indicate the molar percent of each protein. Abbreviations as in Figure 3.
Table II

*N-terminal Amino Acid Sequence of the Branched Intermediate from the G451H Mutant in the Mja Rpol A’ MEIEP Precursor.*

| Cycle | Mja Rpol A’ Intein | MBP          |
|-------|-------------------|--------------|
|       | Residue | pmol | Residue | pmol |
| 1     | S’-PTH | 6.78 | Met     | 12.5 |
| 2     | Val    | 15.2 | Lys     | 6.03 |
| 3     | Asp    | 7.30 | Thr     | 6.95 |
| 4     | Gly    | 5.30 | Glu     | 5.40 |
| 5     | Asp    | 7.59 | Glu     | 7.22 |
| 6     | Thr    | 6.21 | Gly     | 9.46 |
| 7     | Thr    | 7.98 | Lys     | 4.11 |
| 8     | Val    | 12.4 | Leu     | 6.21 |
| 9     | Leu    | 7.23 | Val     | 11.5 |
| 10    | Leu    | 5.36 | Ile     | 2.28 |
| 11    | Asp    | 6.77 | Trp     | 5.31 |
| 12    | Gly    | 5.23 | Ile     | 4.74 |
| 13    | Lys    | 6.23 | Asn     | 8.88 |
| 14    | Leu    | 4.14 | Gly     | 7.59 |
| 15    | Ile    | 2.74 | Asp     | 7.70 |

* Each cycle of Edman degradation contained 2 residues which corresponded to the N-terminal sequence of the intein and the N-extein (MBP).

*S’-PTH* (dehydroalanine PTH artifact) was observed in place of Cys.
Table III

The Effect of Mutating the Mja Rpol A' Intein Penultimate Residue in MEIEP

| Protein | G451 | G451A | G451H | G451K | G451F | G451W |
|---------|------|-------|-------|-------|-------|-------|
| MEIEP*  | 0\(^a\) | 2 ± 1 | 6 ± 2 | 2 ± 1 | 8 ± 1 | 14 ± 1 |
| MEIEP   | 11 ± 1 | 21 ± 7 | 10 ± 3 | 21 ± 8 | 9 ± 1 | 25 ± 7 |
| MEEP    | 68 ± 4 | 51 ± 3 | 50 ± 10 | 50 ± 5 | 49 ± 4 | 28 ± 4 |
| IEP     | 21 ± 4 | 26 ± 5 | 33 ± 6 | 27 ± 3 | 34 ± 4 | 32 ± 4 |
| EP      | 0     | 0     | 0     | 0     | 0     | 0     |

\(^a\) Values indicate the molar percent of each protein. Abbreviations as in Figure 3.
A. Core Endonuclease Domain

B. Intein

| Block A | Block G |
|---------|---------|
| Mja PEP  | CIEGDAKLTLDRG - (412) - TPIVV F N/C |
| Mja Rpol A' | CVDGDVTVLDDLGK - (452) - NGFLT G N/C |
| Ceu ClpP | CLTSDHTVLTTTRG - (456) - KAVWT G N/S |
| CIV RIR1 | CVAPETMILTEDG - (339) - NGILT G Q/C |
| Aae RIR1 | CFIEGTEVLTKRG - (346) - SVCIA G N/C |
| Spb RIR1 | CVTGETLLLTENG - (385) - NGIVT G N/C |
| Ssp DnaE | CLSFGETIILTVEY - (159) - NGAIA A N/C |
| Mja KlbA | ALAYDEPIYLSDG - (168) - EGFAV S N/C |
| Pfu KlbA | ALYDFSVIQLSNG - (523) - NGILV S N/C |
| Pab KlbA | ALYYFSEIIQLPNG - (197) - NGIVV V S N/C |
| Pho KlbA | ALYDFSIIIQLSNG - (519) - NGILV S N/C |
| Pho LHR | CVSGDSKVLTEKRG - (475) - NGFVS K N/S |
| Pho Lon | CFSGEEVIIKEKG - (474) - NGLFV K N/S |
| Pfu Lon | CFSGEEVIIKEKD - (402) - NGLFV K N/S |
| Pab Lon | CFSGEEVIIKRETK - (334) - NGLFV K N/S |
| Pab Moaa | CFPPTEESAVFKFG - (456) - DGILV S N/C |
| Pab RIR1-2 | CFTGDTRILTEKG - (439) - NGIYT S N/S |

83 inteins

C. Mja PEP MEIEP Precursor (135 kDa)

| M | E | Intein |
|---|---|--------|
|  |  | C (150 aa) |
|  |  | (412 aa) |
|  |  | FN (7 aa) |

Mja Rpol A' MEIEP Precursor (145 kDa)

| M | E | Intein |
|---|---|--------|
|  |  | C (200 aa) |
|  |  | (452 aa) |
|  |  | GN (5 aa) |
N-X acyl shift

Transesterification

Asn cyclization

Excised intein

X-N acyl shift

Ligated exteins
Protein splicing in the absence of an intein penultimate histidine
Lixin Chen, Jack Benner and Francine B. Perler

J. Biol. Chem. published online April 18, 2000

Access the most updated version of this article at doi: 10.1074/jbc.M000178200

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