Copyright © 1997 by the Genetics Society of America

Probing Novel Elements for Protein Splicing in the Yeast Vma1 Protozyme: A Study of Replacement Mutagenesis and Intragenic Suppression

Satoru Nogami,* Yoshinori Satow,† Yoshikazu Ohya* and Yasuhiro Anraku*

**Department of Biological Sciences, Graduate School of Science and †Faculty of Pharmaceutical Sciences, University of Tokyo, Tokyo 113, Japan*

Manuscript received March 3, 1997
Accepted for publication May 23, 1997

ABSTRACT

Protein splicing is a compelling chemical reaction in which two proteins are produced posttranslationally from a single precursor polypeptide by excision of the internal protein segment and ligation of the flanking regions. This unique autocatalytic reaction was first discovered in the yeast Vma1p protozyme where the 50-kD site-specific endonuclease (VDE) is excised from the 120-kD precursor containing the N- and C-terminal regions of the catalytic subunit of the vacuolar H^+-ATPase. In this work, we randomized the conserved valine triplet residues three amino acids upstream of the C-terminal splicing junction in the Vma1 protozyme and found that these site-specific random mutations interfere with normal protein splicing to different extents. Intragenic suppressor analysis has revealed that this particular hydrophobic triplet preceding the C-terminal splicing junction genetically interacts with three hydrophobic residues preceding the N-terminal splicing junction. This is the first evidence showing that the N-terminal portion of the V-ATPase subunit is involved in protein splicing. Our genetic evidence is consistent with a structural model that correctly aligns two parallel β-strands ascribed to the triplets. This model delineates spatial interactions between the two conserved regions both residing upstream of the splicing junctions.

PROTEIN splicing is a compelling reaction in which an intervening segment is autocatalytically excised from a nascent polypeptide and the flanking N- and C-terminal regions are ligated with a peptide bond to produce two mature proteins. This splicing reaction was first discovered in a gene of *Saccharomyces cerevisiae* named *VMA1*, which encodes the 70-kD catalytic subunit of the vacuolar H^+-ATPase (referred to as V-ATPase: Hirata *et al.* 1990; Kane *et al.* 1990). The coding region of the 70-kD V-ATPase subunit is interrupted by an in-frame insert that encodes a 50-kD polypeptide (Cooper *et al.* 1993; Kawasaki *et al.* 1996). The excised 50-kD internal segment is a site-specific endonuclease (originally named VDE: *VMA1*-derived endonuclease, Gimble and Thorner 1992). VDE cleaves the specific sequence in a particular allele of *VMA1* that lacks the VDE-coding region (referred to as *VMA1Δvde*), and this site-specific cleavage initiates the meiosis dependent "gene homing" that leads to the insertion of the VDE-coding region into *VMA1Δvde* in a heterozygous strain carrying the *VMA1/VMA1Δvde* allele (Gimble and Thorner 1992).

The primary translational product encoded by the *VMA1* locus is proposed to be called the Vma1 protozyme, named after *protos enzyme*, as it has dual roles for the host organism as a self catalyst for protein splicing and as a selfish element for gene homing (Anraku and Hirata 1994). The names intein and exteins are proposed for the intervening polypeptide to be excised and both the resulting external segments (Perler *et al.* 1994). Since the initial discovery of protein splicing, protozymes have been identified in a number of bacterial, archaeal and eukaryotic organisms covering three major phylogenic trees (Davis *et al.* 1992, 1994; Perler *et al.* 1992; Gu *et al.* 1993). Existence of other putative protozymes has recently been pointed out by examining the DNA sequence database and genome sequencing (Bult *et al.* 1996; Pietrokovski 1994, 1996; Fsihi *et al.* 1996; and for review, see Perler *et al.* 1997).

Amino acid alignment suggests that the coding region of protozymes is a selfish gene that is a "parasite" in the DNA region encoding the external segments. The external segment that serves as the "host" is functionally important. For instance, the external segments encode functionally different enzymes (Figure 1), although most of them share a nucleotide or DNA binding property. Each internal segment contains several conserved motifs (Pietrokovski 1994), especially dodecapeptide motifs that are commonly found in homing endonucleases and are required for endonuclease activity (Lambowitz and Belfort 1993).

Consistent with the idea that the internal segment is a product of the selfish gene, accumulating evidence suggests that the functional domain required for protein splicing is mainly within the internal segment. From a mutational analysis of the Vma1 protozyme, the residues Cys/Ser/Thr at the two splicing junctions and Asn at the C-terminal junction, all of which are con-

Corresponding author: Yasuhiro Anraku, Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan. E-mail: anraku@biol.s.u-tokyo.ac.jp

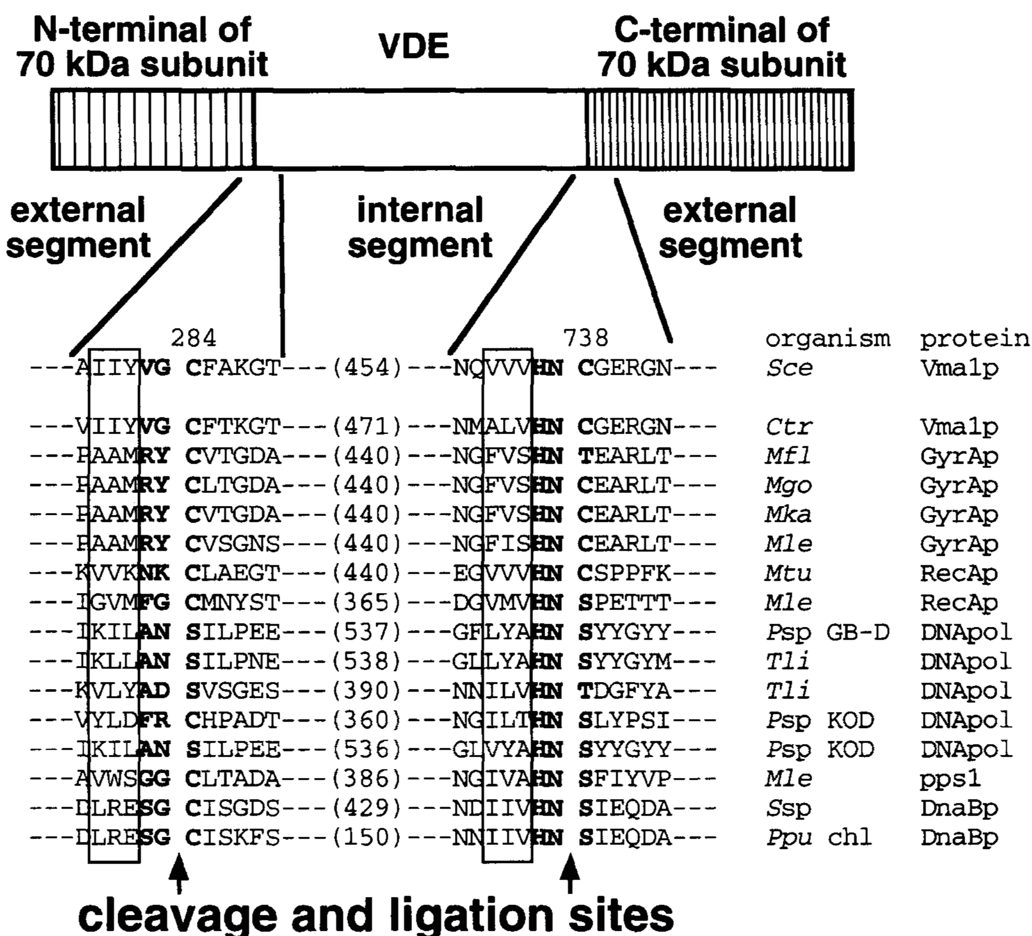


FIGURE 1.—Alignment around the splicing junctions of known or predicted protozymes. In the Vma1 protozyme, the 70-kD subunit of V-ATPase and 50-kD VDE correspond to the external (▥) and internal (□) segments, respectively. Amino acids are shown as one-letter symbol. The number of predicted amino acid residues of the internal segments is indicated in parentheses. Highly conserved residues are shown in bold, and conserved hydrophobic residues are boxed. Proteins encoded in the external segments are described in the right columns. Arrows indicate the reaction sites of protein splicing. *Sce, Saccharomyces cerevisiae; Ctr, Candida tropicalis; Mfl, Mycobacterium flavescens; Mgo, Mycobacterium gordonae; Mka, Mycobacterium kansasii; Mle, Mycobacterium leprae; Mtu, Mycobacterium tuberculosis; Psp, Pyrococcus* species; *Tli, Thermococcus litoralis; Ssp, Synechocystis* species; *Ppu* chl, *Porphyra purpurea* chloroplast; DNApol, DNA polymerase.

served among protozymes (Figure 1, shown in bold) have been shown to be critical for the reaction (HIRATA and ANRAKU 1992; COOPER *et al.* 1993). The internal segments were reported to be excised from an unrelated insertional context *in vivo* in yeast (COOPER *et al.* 1993) and *in vitro* in an extreme thermophilic archaeal protozyme (XU *et al.* 1993, 1994). Moreover, folding-dependent *in vitro* protein splicing of the yeast Vma1 protozyme has revealed that VDE bracketed by only six proximal and four distal amino acids is processed autocatalytically (KAWASAKI *et al.* 1996).

In this report, we have probed novel elements for protein splicing by a systematic genetic study. By applying a random replacement mutagenesis technique (PALZKILL and BOTSTEIN 1992), we have demonstrated that a hydrophobic triplet, the residues $Val_{733}Val_{734}$-Val_{735} located in front of the highly conserved His-Asn-Cys/Ser/Thr at the C-terminal junction, is required for protein splicing (The numbering of the amino acid sequence in this report refers to the numbering of the original *VMA1* gene product; HIRATA *et al.* 1990). This unique valine triplet was then shown to interact genetically with a hydrophobic amino acid triplet, residues $Ile_{279}Ile_{280}Tyr_{281}$, which precede the N-terminal junction site. On the basis of these results, we discuss a possible functional arrangement of the structural elements for protein splicing catalysis in the Vma1 protozyme.

MATERIALS AND METHODS

Terms used in this study: Vma1 protozyme means the 120-kD primary translation product of the *VMA1* gene. *VMA1* is identical to the *TFP1* gene (HIRATA *et al.* 1990) and the original sequence of *TFP1-408* was described by SHIH *et al.* (1988). Vma1p and VDE indicate the 70-kD subunit of V-ATPase and the 50-kD endonuclease, respectively, both of which are produced from the Vma1 protozyme by protein splicing. External segments are the part of the Vma1 protozyme composing Vma1p and the internal segment is that composing VDE. Vma1 protozyme, Vma1p, VDE, the external segments and

the internal segment are also called precursor protein, mature protein, PI-SceI, exteins and intein, respectively (PERLER *et al.* 1994).

Yeast media and strains: Yeast media and genetic techniques were described in KAISER *et al.* (1994). Cells were grown in YPD medium or YNBD medium with casamino acids (Difco) and appropriate supplements. Ca^{2+}-sensitivity of cells was examined on YPD plates supplemented with 100 mM $CaCl_2$ (OHYA *et al.* 1991). Medium pH was adjusted by adding 50 mM succinate/NaOH buffer. A *S. cerevisiae* strain NY101 was created by disrupting the *VMA1* locus of YPH499 (*MAT***a** *ade2 his3 leu2 lys2 trp1 ura3*; SIKORSKI and HIETER 1989) by replacing a 2.8-kb *Stu*I-*Sph*I region of the *VMA1* gene with a 1.1-kb *Sma*I-*Sph*I fragment of pJJ242 (JONES and PRAKASH 1990), which contains the *URA3* gene. NY101 did not grow on YPD medium supplemented with 100 mM $CaCl_2$ (Cls^- phenotype), which was one of the characteristic phenotypes of the V-ATPase deficient strains (OHYA *et al.* 1991). In addition, NY101, which harbors the *ade2* mutation, failed to accumulate a pigment derived from the adenine biosynthetic intermediates due to the deletion mutation of *vma1*. NY101 formed white colonies on YNBD supplemented with casamino acids and buffered at pH 5.0 while YPH499 formed red colonies.

DNA manipulations: DNA manipulations were performed using standard methods (SAMBROOK *et al.* 1989). Pfu polymerase and calf intestine phosphatase were purchased from Stratagene and Boehringer-Mannheim, respectively. All the other enzymes were obtained from TAKARA and New England Biolabs. Oligonucleotides were synthesized using DNA synthesizer 394 (ABI) or purchased from Greiner Japan. Purification of DNA fragments was routinely carried out by electrophoresis in agarose followed by treatment with DNA prep (Dia Iatron). DNA Engine (MJ Research) was used to amplify DNA fragments. DNA sequencing was performed using an autosequencer 373A (ABI).

Plasmids: Plasmid pET17b-VDEWT (KAWASAKI *et al.* 1997) and its derivatives were used for expressing mutant VDEs in *Escherichia coli*. Plasmid pBSΔVDE was used as substrate for the endonuclease assay (KAWASAKI *et al.* 1997). Plasmid pYO314 was an autonomous, yeast centromere-containing vector marked with *TRP1* (Y. OHYA, unpublished). It contains the amp^r gene that is resistant to digestion with *Bsa*I. The *Bam*HI-*Sal*I fragment of pMVMA1 (HIRATA and ANRAKU 1992), which contains the *VMA1* gene, was subcloned into pYO314 to generate pSN001. A *Bsa*I site on pSN001 (193 bp upstream of the translation initiation site of *VMA1*) was disrupted using an oligonucleotide 5′-GGC CAC GGC CGC TCA AAG GGG ACT CAC ACT GGA TTA AAA-3′ to create pSN101. This mutant allele fully complemented the *vma1* disruption, because yeast NY101 carrying pSN101 was able to grow on calcium-rich medium and to form red colonies on YNBD medium with casamino acids buffered at pH 5.0. All the other plasmids used in this study were derivatives of plasmid pSN101.

Random replacement mutagenesis: Three codons at positions 733–735 of the *VMA1* gene were replaced with random sequences using random replacement mutagenesis (PALZKILL and BOTSTEIN 1992) with some modifications. The protocol adopted was as follows: The nucleotide sequence 5′-GTT GTC GT-3′ at positions 2197–2204 (covering the codons 733–735) of *VMA1* was replaced by 5′-GAG ACC ACT AGT GGT CTC 3′ by a two-step PCR method (HO *et al.* 1989). As a result, the original sequence at the region of mutagenesis was replaced by oligonucleotides containing *Bsa*I recognition sites (underlined). The inserted oligonucleotide was released by digestion with *Bsa*I, which created a 10-bp deletion. To replace the deleted nucleotides with random sequences, a second linker (made from oligonucleotides 5′-NNN NNC GAG ACC CTC GAG GCT AGC GGT-3′ and 5′-NNN NNC GAG ACC GCT AGC CTC GAG GGT-3′ by annealing and blunting) was inserted that contains 5 bp of random sequence at each end along with embedded *Bsa*I recognition sites. Libraries of independent linker insertions were constructed in *E. coli*, and the plasmid DNA was extracted and purified. The DNA was digested with *Bsa*I again and religated, leaving an insertion of 10 random nucleotides.

Two independent plasmid libraries containing randomly substituted *VMA1* mutations at codons 733–735 were used in this study. Plasmids were introduced into yeast strain NY101 harboring the *vma1* deletion mutation. Transformants formed colonies with various colors from red to white, according to the complementing abilities of the randomly substituted *VMA1* mutations. We chose transformants that formed red or pink colonies at 30° to select mutants that were expected to have functional *VMA1* mutations. Plasmids containing functional *VMA1* mutations were recovered from yeast, amplified in *E. coli*, and transformed back into NY101 to ensure that the phenotype was conferred by the plasmid rather than by any spontaneous genomic mutations. DNA sequencing of random replacements at codons 733–735 revealed that some mutants had the same amino acid sequence as other mutants. To avoid duplication, we counted one when the same nucleotide sequence was obtained from the same plasmid pool and counted two when the same DNA sequences were obtained from different pools or when different DNA sequences encoding the same amino acid were obtained.

For site-directed suppressor analysis, random replacement mutagenesis was used to randomize three codons at positions 279–281 of the seven splicing mutants carrying mutations at the positions 733–735 (SerLeuGln from group 2, GlyLeuVal and ValLysSer from group 3 and GlyLeuTyr, LeuLeuLeu, SerGlyIle and ValMetAsp from group 4 mutants). Plasmid DNAs of these suppressor mutants were recovered, rescued in *E. coli*, reintroduced into yeast to examine reproducibility of the phenotypes and prepared for DNA sequencing.

Randomization using error-prone PCR: Suppressor mutations in the *VMA1* gene were created by the PCR mutagenesis method (MUHLRAD *et al.* 1992). The procedure was first to amplify the region of interest of *VMA1* under mutagenic PCR conditions (50 mM KCl, 10 mM Tris-Cl (pH 8.3), 0.01% gelatin, 7 mM $MgCl_2$, 0–5 mM $MnCl_2$, 0.2 mM each dATP and dGTP, and 1.0 mM each dCTP and dTTP; CADWELL and JOYCE 1992) using primers 5′-TAC TTG GCC TGT TCG TGT TCC 3′ and 5′-TGG GAA TTC CAT CAA GAC TTC TGC CAT TTC 3′. After amplification, PCR products were digested with *Bcl*I and gel-purified. The resultant DNA fragment corresponds to amino acid positions 214–724 of the *VMA1* gene. Then, plasmids harboring *vma1* mutations (carrying amino acid substitutions LeuLeuLeu, SerLeuGln, GlyLeuVal or ValLysSer at codons 733–735 that confer a protein splicing defect) were digested with *Hpa*I and *Sac*II to create a 1405-bp gap in the VDE-coding region of *VMA1* (corresponds to amino acid position from 234 to 700). The gapped plasmids contain homology to both ends of the mutagenized PCR product. Each gel-purified gapped plasmid was cotransformed with the mutagenized PCR product. The transformants were plated on YPD plates supplemented with 100 mM $CaCl_2$ and incubated at 30° (for the LeuLeuLeu mutant) and 37 (for the SerLeuGln, GlyLeuVal or ValLysSer mutants) for selecting suppressor mutants. Plasmid DNAs of these suppressor mutants were recovered, rescued in *E. coli*, reintroduced into yeast to confirm the phenotypes and prepared for DNA sequencing.

To identify the mutation responsible for the suppression, the subcloned *Bgl*II-*Bgl*II fragment or the *Bam*HI-*Kpn*I fragment of the suppressor *VMA1* gene was replaced with the original splicing-deficient *vma1* gene.

Antibodies and Western blotting: Anti-70-kD Vma1p monoclonal (R70, HIRATA and ANRAKU 1992) and anti-VDE polyclonal antibodies (KAWASAKI *et al.* 1996), which recognize the N-terminal region of the 70-kD subunit of V-ATPase and 50-kD VDE, respectively, were used for analyzing plasmid-born *VMA1* products by Western blotting analysis. Preparation of cell extracts and Western blotting analysis were performed as described previously (HIRATA *et al.* 1990).

Purification and endonuclease activity assay of mutant VDEs: For expressing mutant VDEs in *E. coli*, the 1.4-kb *Kpn*I-*Eco*RI fragment of the mutant *VMA1* gene was replaced with the same fragment of pET17b-VDEWT (KAWASAKI *et al.* 1997). Expression, purification of VDEs and measurement of their endonuclease activities were as described elsewhere (KAWASAKI *et al.* 1997).

RESULTS

Substitutions at amino acid positions 733–735 of *VMA1* confer various abilities to complement *vma1*: A hydrophobic amino acid cluster upstream of the C-terminal splicing junction is conserved among protozymes (Figure 1). To investigate the requirement of this conserved region for protein splicing, we replaced the valine triplet with randomized sequences at amino acid positions 733–735 of the Vma1 protozyme using a random replacement mutagenesis technique (PALZKILL and BOTSTEIN 1992). This method enables us to randomize the DNA sequence of a short stretch of a gene completely and to determine the percentage of all possible random sequences that produce a functional protein.

Plasmids encoding "randomized" Vma1 protozymes were introduced into yeast strain NY101 harboring a deletion mutation of *vma1*. As described in MATERIALS AND METHODS, transformants formed colonies with various colors from red to white. To select mutants that were expected to have functional *VMA1* mutations, we therefore chose transformants that formed red or pink colonies at 30°. The ability of the mutants to complement the *vma1* disruption mutation was further examined using their Cls$^-$ phenotypes at 23, 30 and 37°, which are characteristic of V-ATPase defective mutants (OHYA *et al.* 1991). Several randomized mutants showed the Cls$^-$ phenotype at 37°, but not at 30 or 23° (for examples at 23 and 37°, see Figure 2A). On the basis of the Ca^{2+} sensitivity at 30 and 37°, we classified the functional mutants into three groups. Group 1 (26 mutants) grew well at any temperature examined, group 2 (20 mutants) grew well at 30° but did not grow or grew very slowly at 37°, group 3 (20 mutants) grew slowly at 30° but did not grow at 37°. Once correctly spliced from the mutated 120-kD Vma1 protozyme, the 70-kD Vma1p is expected to be as functional as the wild-type V-ATPase subunit, since random mutations are introduced only in the VDE-coding region. The inability of *VMA1* mutations to complement the *vma1* disruption is therefore ascribable to loss of the protein splicing reaction.

We selected yeast transformants that formed white colonies as a control. These strains showed Cls$^-$ phenotypes at any temperature examined and they were classified as group 4 (15 mutants). As discussed later, however, 11 of them were inappropriate for further analysis because they carried nonsense or frame-shift mutations at the mutated region.

Defects of protein splicing result in inability to complement the *vma1* disruption: Defective protein splicing by the randomized mutants was examined by Western blotting analysis (Figure 2B). In the lysate prepared from wild-type cells, the 50-kD VDE protein and the 70-kD subunit of V-ATPase were detected with an anti-VDE antibody and an anti-Vma1p antibody, respectively (Figure 2B, lanes 1 and 12). Randomized mutants showed a spectrum of phenotypes ranging from no obvious inhibition (group 1 mutants; Figure 2B, lanes 2–3 and 13–14) to a complete blockage of splicing (group 4 mutants; Figure 2B, lanes 10 and 11), consistent with the result of their Cls$^-$ phenotypes. The splicing reaction was temperature dependent in several mutants that showed the Cls$^-$ phenotype only at 37°. In these mutants, the ratio of the 120-kD precursors to mature products (50 kD VDE or 70 kD Vma1p) was greater at 37° than at 23°. In addition, neither processed products nor byproducts were seen in cells of noncomplementing mutants (group 4), resulting in accumulation of 120-kD mutant Vma1 protozymes (Figure 2B, lanes 10 and 11). These results indicated that amino acid substitutions at positions 733–735 of *VMA1* crucially affect protein splicing.

Hydrophobic or uncharged amino acids upstream of the C-terminal junction are required for protein splicing: DNA sequences of the randomized mutants were determined to identify the allowable substitutions at positions 733–735. After omitting duplicates (see MATERIALS AND METHODS), the deduced amino acid sequences of 62 functional mutants are listed in Table 1. Amino acid sequences of four missense mutants of 15 group 4 mutants that did not grow on calcium-rich medium are also listed. There was a statistically significant preference for amino acids such as Ala, Cys, Ile, Ser, Thr and Val at any of three positions (Table 1) in the functional mutants. For example, we found that only five amino acids (Ala, Gln, Ser, Thr and Val) appeared at the position 735 of group 1 mutants (Table 1). The probability that this preference of codons encoding these five residues would occur by random chance is $<10^{-6}$ ($\chi^2 = 51$ with 1 d.f.). Codons 733 and 734 of group 1 mutants showed similar preferences (data not shown). Moreover, charged amino acids were not allowed in group 1 mutants ($P < 0.01$ for each position; $\chi^2 = 7$ with 1 d.f.). Independently isolated mutants carrying the same amino acid alteration (for example ValLeuAla mutations in group 1; Table 1) also suggested a limited allowance of amino acids in this region. This preference of amino acids at the triplet 733–735 indicated that hydrophobic or uncharged side chains are required for the splicing reaction.

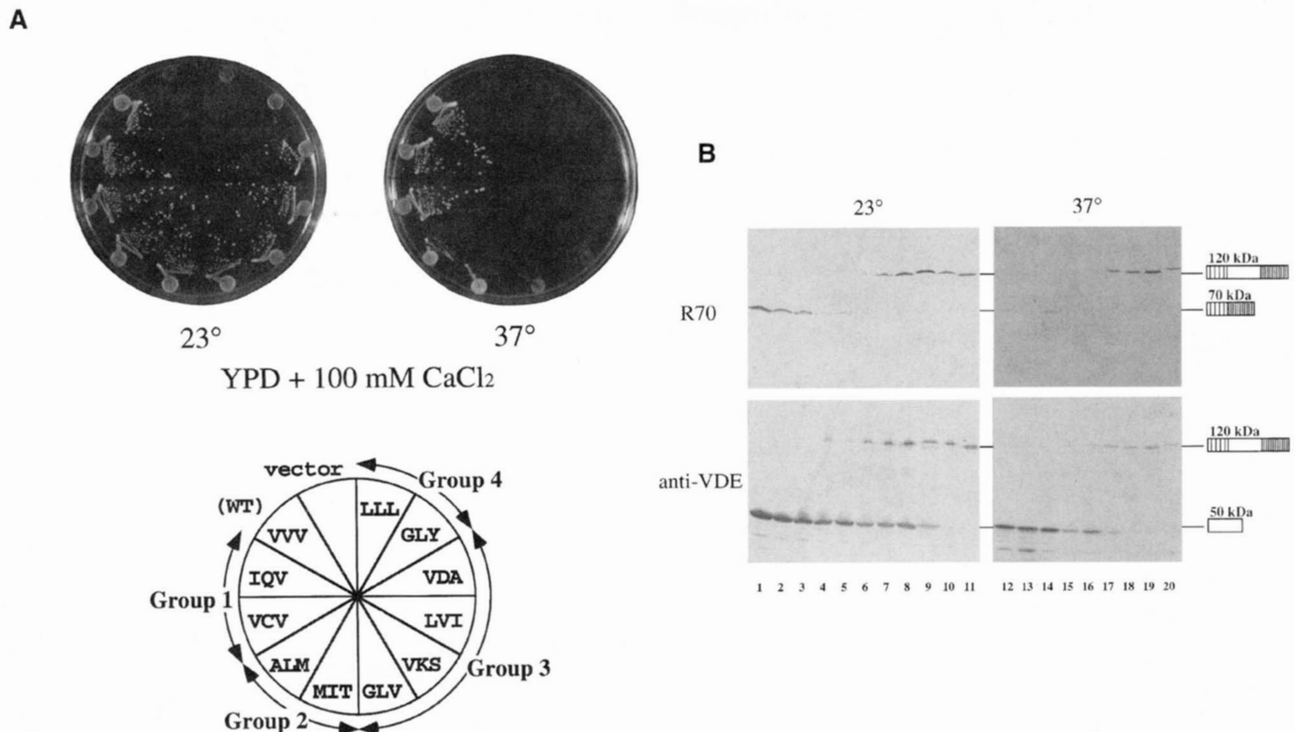


FIGURE 2.—Replacement of the hydrophobic triplet residues proximal to the C-terminal splicing junction of the Vma1 protozyme affects the splicing reaction. (A) Growth on calcium-rich YPD plates of strain NY101 that was transformed with plasmid pYO314 (vector), pSN101 (WT), or pSN101 carrying the substitutions (one-letter symbol) indicated at positions 733–735 of the *VMA1* gene product. Cells were grown at the indicated temperatures on YPD medium supplemented with 100 mM $CaCl_2$. Classification of mutants is described in RESULTS. (B) Strain NY101 was transformed with plasmid pSN101 carrying the following substitutions at positions 733–735 of the *VMA1* gene product: VVV (wild-type; lanes 1 and 12), IQV (lanes 2 and 13), VCV (lanes 3 and 14), ALM (lanes 4 and 15), MIT (lanes 5 and 16), GLV (lanes 6 and 17), VKS (lanes 7 and 18), LVI (lanes 8 and 19), VDA (lanes 9 and 20), GLY (lane 10) and LLL (lane 11). Cell extracts were prepared from cells cultivated in YNBD medium supplemented casamino acids and buffered at pH 5.0 at 23° (lanes 1–11) or 37° (lanes 12–20), resolved by SDS-PAGE (7.5% acrylamide gel) and used in Western blotting analysis probed with either an anti-Vma1p monoclonal antibody (R70: top) or an anti-VDE polyclonal antibody (anti-VDE: bottom). ▥ and □, external and internal segments, respectively.

Of course, hydrophobicity is not the only issue since the LeuLeuLeu mutation is nonfunctional (Table 1, group 4). One of the possible factor that affects the splicing reaction is the volume of side chains. At positions 733 and 735, few of the side chains with volumes larger than Val, such as Trp, Arg, Tyr, Phe, Lys, Leu, Ile, Met, His and Gln, appeared in group 1 mutants (of these, only a glutamine residue was found at position 735). The probability is $<10^{-5}$ at position 735 ($\chi^2 = 16$ with 1 d.f.). However, at position 734, Leu and Ile residues that are larger than Val, were frequently found in the functional mutants (Table 1, group 1). This difference of amino acid preferences suggests that the volume of the side chain is important for the splicing reaction at positions 733 and 735, while it is not important at position 734.

In summary, hydrophobic and uncharged side chains at positions 733–735 are required for proper splicing and, in addition, small side chains are preferred at positions 733 and 735.

Endonuclease activity of defective mutant VDEs for protein splicing: Four protein splicing mutant VDEs carrying mutations at the positions 733–735 (ValValVal to LeuLeuLeu from group 4, GlyLeuVal and ValLysSer from group 3, and SerLeuGln from group 2) were expressed in *E. coli*. The constructs for expressing VDEs produce recombinant VDEs bracketed by 18 proximal and 17 distal amino acids. When expressed at 37°, all constructs other than wild-type produced unspliced VDEs that were recovered in the insoluble fraction (data not shown). When expressed at 23°, however, all constructs but the one that carried the ValValVal to LeuLeuLeu mutations produced spliced VDEs in the soluble fraction. Expressed VDEs in the soluble fraction of *E. coli* lysates were purified by means of cation exchange chromatography.

Purified wild-type VDE cleaved a substrate DNA carrying the recognition sequence of VDE (Figure 3B, lane 2). Mutant VDEs carrying ValValVal to SerLeuGln, GlyLeuVal and ValLysSer mutations at 733–735 were able to cleave the same substrate (Figure 3B, lane 3–5) as efficiently as wild-type VDE (Figure 3B, lane 2). This result indicated that mutations at 733–735, which cause defects in protein splicing, retain specificity and activity for the endonuclease, suggesting that the global conformation correctly retained in the mutant VDEs.

TABLE 1

Amino acid sequence at codons 73–735 of the mutant *VMA1*

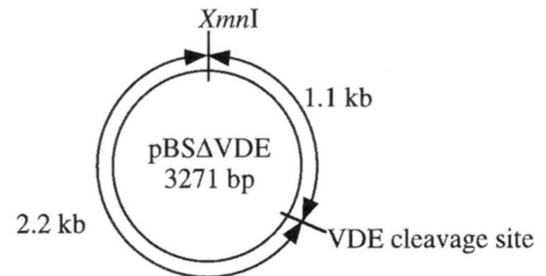
Group 1	Group 2	Group 3	Group 4
Ala Val Thr	Ala His Val	Cys Thr Ala	Gly Leu Tyr
Cys Ile Thr	Ala Ile Thr	Gly Leu Val	Leu Leu Leu
Cys Leu Thr	Ala Leu Met	Gly Val Ala	Ser Gly Ile
Gly Val Val	Ala Leu Thr	Leu Leu Met	Val Met Asp
Ile Gln Val	Asn Leu Val	Leu Ser Val	
Ile Leu Ala	Cys Glu Val	Leu Val Ile	
Ile Leu Ala	Cys Val Leu	Leu Val Ile	
Ile Leu Thr	Leu Leu Ser	Ser Glu Val	
Ile Pro Val	Leu Leu Thr	Ser Leu Ala	
Ile Ser Val	Met Ile Thr	Thr Leu Leu	
Ile Ser Val	Pro Ser Val	Val Asp Ala	
Met Val Val	Ser Leu Gln	Val Glu Ala	
Phe Ile Val	Val Ala Thr	Val Leu Gln	
Thr Val Ala	Val Ala Thr	Val Leu Met	
Thr Val Gln	Val Arg Ile	Val Lys Ile	
Thr Val Thr	Val Arg Thr	Val Lys Ser	
Val Cys Val	Val Asn Thr	Val Tyr Ala	
Val Ile Val	Val Asn Thr		
Val Leu Ala	Val Ile Gln		
Val Leu Ala	Val Pro Val		
Val Leu Thr			
Val Val Ser			
Val Val Thr			
Val Val Thr			
Val Val Val			

Mutants are classified as described in RESULTS. In Group 4 mutants, truncated and frame-shift mutants are not shown. Amino acids that did not appear in functional mutants (Group 1–3) are; Arg, Asp, Gln, Glu, His, Lys, Trp and Tyr at position 733, Gly, Met, Phe and Trp at position 734, and Arg, Asn, Asp, Cys, Glu, Gly, His, Lys, Phe, Pro, Trp and Tyr at position 735.

Identification of the regions that ensure the splicing reaction by intragenic suppression: Hydrophobic interaction with other residues may be a plausible explanation for the functional involvement of the valine triplet in the splicing reaction. To identify putative residues that functionally interact with the triplet valines in the Vma1 protozyme, an intragenic suppressor study was systematically carried out.

By PCR mutagenesis (CADWELL and JOYCE 1992), we introduced random mutations in the region (see Figure 4) of four splicing mutants, three of which were shown to have no defect of endonuclease activity (see Figure 3). This mutagenized region, covering a wide region from position 214 to 724 of the Vma1 protozyme, contains residues preceding the N-terminal junction and most of the VDE region. To avoid a reversion mutation, we excluded the positions 733–735 from mutagenesis. Plasmids carrying an intragenic suppressor mutation should give rise to Cls$^+$ transformants at temperatures at which mutant parents could not grow. Cls$^+$ transformants were therefore selected on calcium-rich medium at 37° (for group 2 and 3) and at 30° (for group

A



B

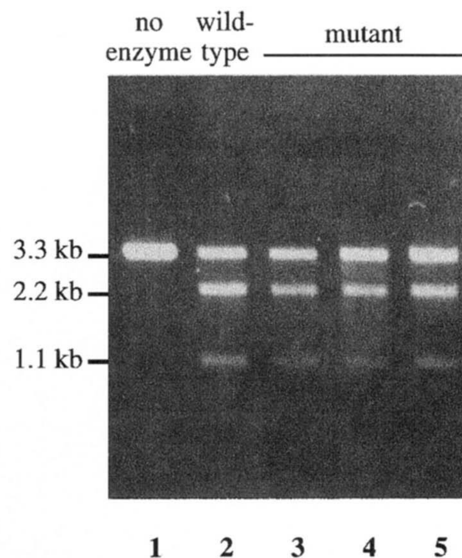


FIGURE 3.—Endonuclease activity of the protein splicing mutant VDEs. (A) Schematic presentation of plasmid pBSDVDE used as DNA substrate in this experiment. The unique *Xmn*I site in the vector and the VDE-recognition site are indicated. The predicted sizes of the two DNA fragments produced by VDE cleavage from the plasmid that was linearized by digestion with *Xmn*I are also shown. (B) pBSDVDE linearized with *Xmn*I was incubated at 25° for 60 min with no enzyme (lane 1) or purified VDEs carrying the following amino acid substitutions (one-letter symbol) at positions 733–735: VVV (wild-type, lane 2), SLQ (lane 3), GLV (lane 4), VKS (lane 5). After incubation, DNA was extracted with phenol/chloroform and analyzed by electrophoresis in a 0.7% agarose gel.

4). The numbers of suppressor mutants obtained were four from the SerLeuGln mutant, two from the GlyLeu-Val mutant, and two from the ValLysSer mutant. No suppressor mutant was obtained from the LeuLeuLeu mutant.

Mutation points of eight suppressors were determined by DNA sequencing of the region corresponding to amino acid positions 210–756 covering the mutated region. We found that seven of the eight mutants carried mutations near the N-terminal junction, although many of them had multiple mutations (Table 2). We

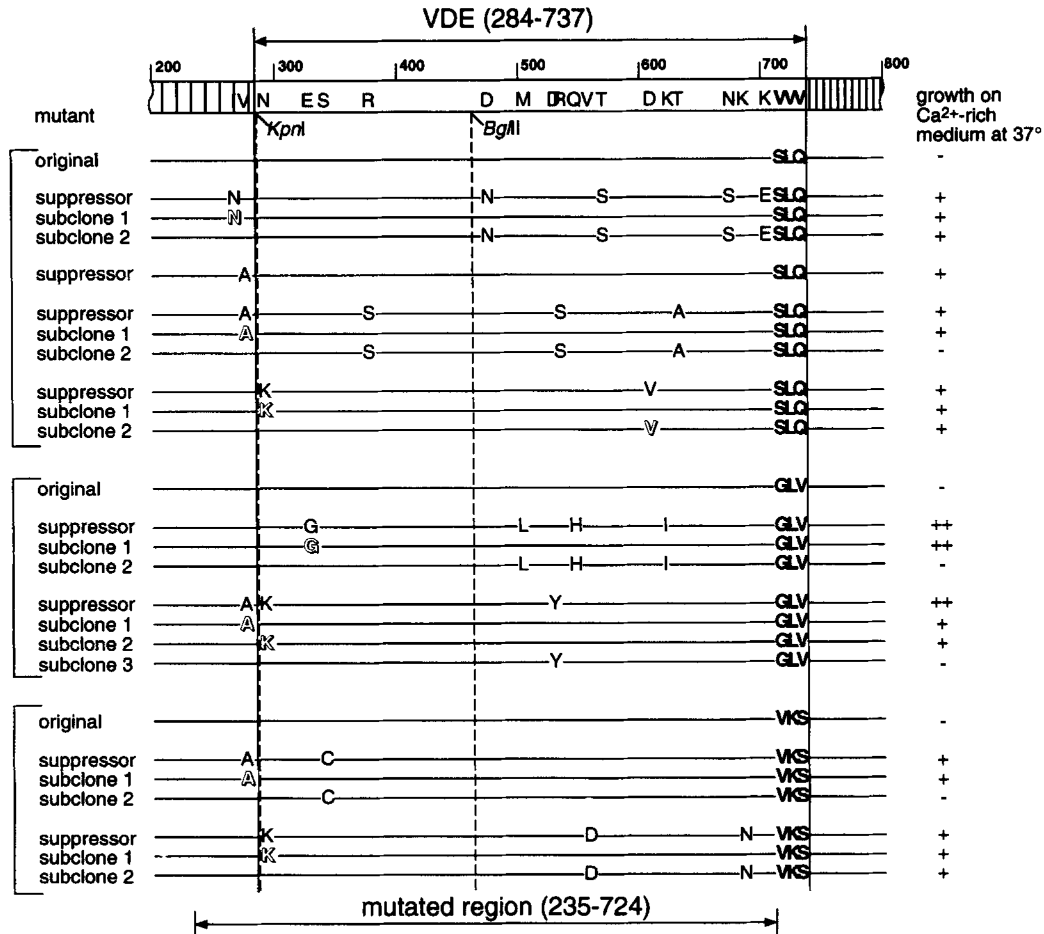


FIGURE 4.—Intragenic suppression mutations (one-letter symbol) of the splicing mutants. ▥ and □, external and internal segments, respectively. Numbers indicate amino acid positions of the Vma1 protozyme. Substituted amino acids in original mutants, suppressor mutants and subclones of suppressor mutants are depicted on horizontal lines. Bold letters represent amino acid substitutions at positions 733–735 that cause the splicing defect. Single amino acid substitutions of the subclones that suppress the splicing defects are shown in outlined letter. The original amino acids in the respective positions of the wild-type Vma1 protozyme are presented in the upper box. The *Kpn*I site and *Bgl*II site used for subcloning of suppressor mutations are also shown as broken lines. In the right-most column, growth of cells on YPD + 100 mM $CaCl_2$ plates at 37° are indicated as follows: ++, equivalent growth to wild-type; +, weak growth; −, no growth.

next determined the residues that caused intragenic suppression by subcloning each mutation. Single amino acid substitutions located near the N-terminal junction ($Ile_{279}Asn$, $Val_{282}Ala$ and $Asn_{290}Lys$) suppressed the growth defect of each SerLeuGln, GlyLeuVal, and ValLysSer mutation at positions 733–735 (Figure 4 and data not shown). The sequences for these suppressors and their derivatives are summarized in Figure 4.

The mutant bearing the triplet GlyLeuVal at positions 733–735 of the *VMA1* gene product failed to grow on calcium-rich medium at 37° (Figure 5A). The $Val_{282}Ala$ mutation and the $Asn_{290}Lys$ mutation each suppressed the Cls^- phenotype caused by the GlyLeuVal mutation (Figure 5), while the $Asp_{537}Tyr$ mutation did not. Suppression of the splicing defect was confirmed by Western blotting analysis. In yeast cells carrying the GlyLeuVal mutation cultivated at 37°, the 120-kD precursor accumulated (Figure 5B, lane 2). In the suppressed cells, however, the amount of the 120-kD Vma1 protozyme was decreased while the amount of the 50 kD VDE protein was partially restored (Figure 5B, lane 3–5). Consistent with the phenotypic properties, the $Val_{282}Ala$ and the $Asn_{290}Lys$ mutations each suppressed the GlyLeuVal mutation (Figure 5B, lane 4 and 5). The Asp_{537}-Tyr mutation did not alter the proportion of the products (Figure 5B, compare lanes 2 and 6), suggesting that this mutation had no effect in the splicing reaction of the GlyLeuVal mutant.

Demonstration of functional interaction between residues preceding the N- and C-terminal junctions by position-specific intragenic suppression: The mutations located in the external segment of VDE ($Ile_{279}Asn$ and $Val_{282}Ala$) suppressed the splicing defect caused by mutations located at the internal segment (Figure 4). This finding suggests that the residues preceding the N-terminal junction may be involved in the protein splicing reaction, particularly through functional interaction with the residues preceding the C-terminal junction.

TABLE 2

Mutation point of intragenic suppressor mutant

Amino acid sequence at positions 733–735	Second-site mutation	Growth on Ca^{2+}-rich medium at 37°
VVV	None	++ (wild-type)
SLQ	None	−
SLQ	I279N, D474N, T565S, N676S, K710E	+
SLQ	V282A	+
SLQ	V282A, R373S, R538S, T621A	+
SLQ	N290K, D609V	+
GLV	None	−
GLV	E328G, M519L, Q542H, K619I	++
GLV	V282A, N290K, D537Y	++
VKS	None	−
VKS	V282A, S332C	+
VKS	N290K, V553D, K685N	+

Amino acid substitutions at positions 733–735 are also listed. In the rightmost column, growth of cells on calcium-rich YPD plates at 37° is indicated as follows: ++, equivalent growth to wild-type; +, weak growth; −, no growth.

Indeed, there are clusters of hydrophobic amino acids in front of the N-terminal junction, which are largely conserved among protozymes, although these segments are the N-terminal half of different enzymes (Figure 1, left box). To test the functional involvement of this new element in the splicing reaction, we randomized the IleIleTyr sequence at positions 279–281 in the seven splicing mutants and tried to isolate site-specific suppressor mutants. Randomized plasmid pools were introduced into the NY101 strain, and cells that grew on YPD medium supplemented with 100 mM $CaCl_2$ at 37° (for group 2 or group 3 mutants) or 30° (for group 4 mutants) were screened. No suppressor mutant was obtained from group 4 mutants (GlyLeuTyr, LeuLeuLeu, SerGlyIle and ValMetAsp at positions 733–735). However, we found that transformants derived from the partial splicing mutants (SerLeuGln, GlyLeuVal and ValLysSer at the positions 733–735) grew better than the original mutants. Table 3 shows the list of 25 sequences at positions 279–281 that suppressed the splicing defect caused by mutations at positions 733–735. Among these mutants, we analyzed one mutant further.

The second-site mutation (AsnValSer at positions 279–281) partially suppressed the Cls^- phenotype at 37° of the ValLysSer mutation at the positions 733–735 (Figure 6A). The AsnValSer mutation alone showed no growth defect when compared with the wild-type strain (data not shown). Suppression of the splicing defect was confirmed by Western blotting analysis. In yeast cells carrying the ValLysSer mutation cultivated at 37°, the 120-kD precursor accumulated (Figure 6B, lane 3). In the suppressed cells, however, the amount of the 120-kD Vma1 protozyme was decreased while the amount of the 50-kD VDE protein was restored (Figure 6B, lane 4).

DISCUSSION

In this study, we have accomplished two major advances in our knowledge of protein splicing. First, we have demonstrated the importance of a conserved hydrophobic cluster residues, in this case $Val_{733}Val_{734}Val_{735}$, which resides in front of the C-terminal junction. Second, we have shown a possible functional interaction of this cluster with a hydrophobic cluster, the $Ile_{279}Ile_{280}Tyr_{281}$ residues that are part of the mature 70-kD subunit of V-ATPase, which is located at the N-terminal junction site. This is the first evidence suggesting that important residues for protein splicing are located not only in the regions encoded by the parasite DNA but also in that encoded by the host DNA.

Conserved hydrophobic residues are required for the reaction: The involvement of the conserved hydrophobic triplet valine residues was established by isolation of mutants carrying substitutions at the positions 733–735 of the Vma1 protozyme that conferred various protein splicing abilities that account well for their growth phenotypes (Figure 2). Western blotting analysis also suggested that the main defect of these splicing mutants was owing to reduced rates of the splicing reaction, rather than side reactions, as revealed by the absence of partial splicing products such as the 90-kD (produced by a cleavage at the N-terminal junction) and the 80-kD (produced by a cleavage at the C-terminal junction) polypeptides (Figure 2B). Thus, in recognition of a previous result (COOPER et al. 1993), we propose that hydrophobic or uncharged amino acids at the positions 733–735 are required for the correct splicing reaction (Table 1).

Evidence for the functional interaction between the positions 279–281 and 733–735: The possible interac-

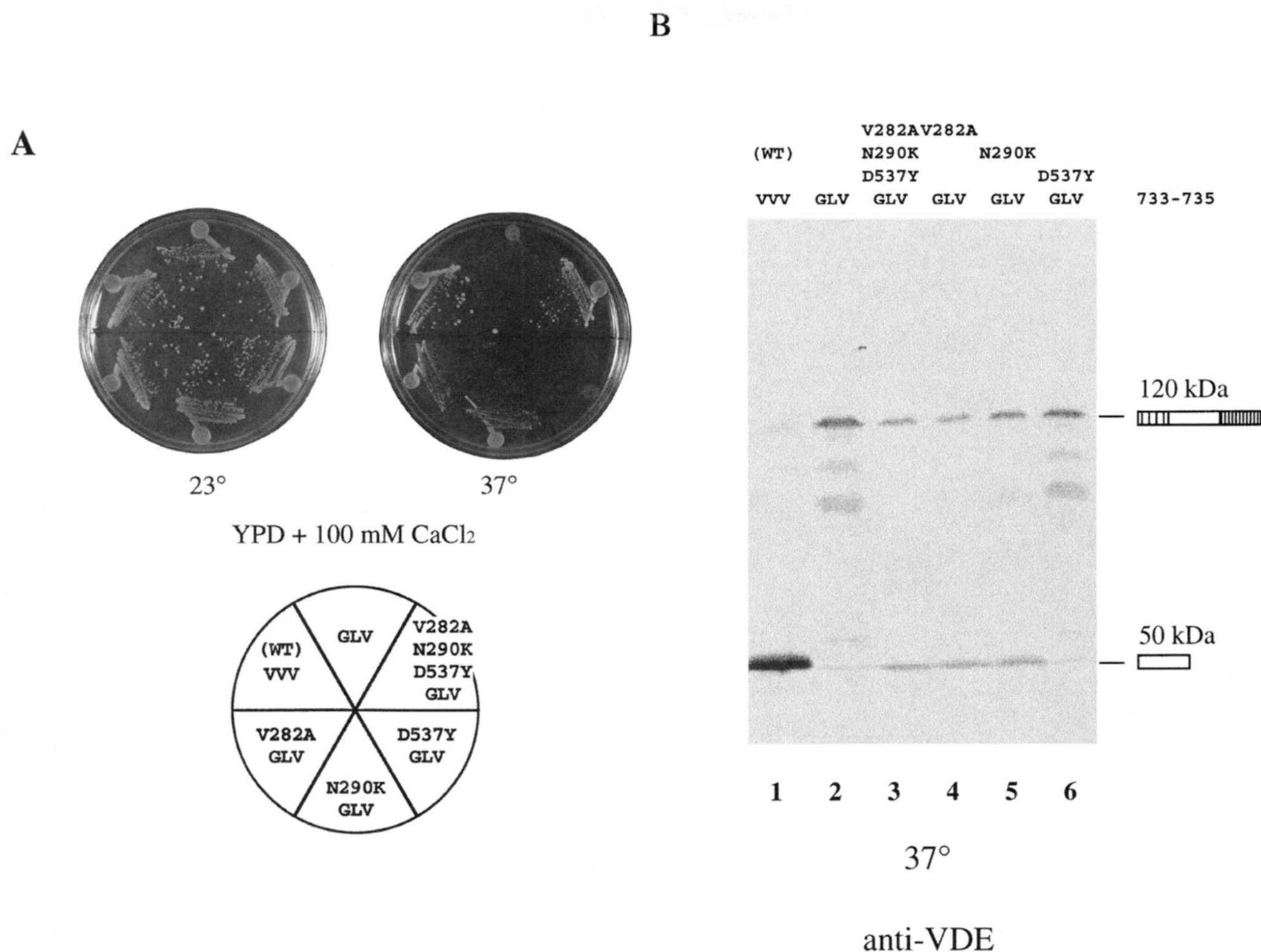


FIGURE 5.—Intragenic suppression of the $Gly_{733}Leu_{734}Val_{735}$ mutant. (A) Growth on calcium-rich YPD plates of strain NY101 that was transformed with the plasmid pSN101 (WT) and pSN101 carrying substitutions (one-letter symbol) at the indicated position and at the positions 733–735 (lowest lines) of the *VMA1* gene product. Cells were grown at the indicated temperatures on YPD medium supplemented with 100 mM $CaCl_2$. (B) Strain NY101 was transformed with the plasmids pSN101 (WT) and pSN101 carrying the substitutions at the indicated positions and at the positions 733–735 (lowest lines) of the *VMA1* gene product. Cell extracts, which were prepared from cells cultured in YNBD medium supplemented with casamino acids and buffered at pH 5.0 at 37°, were resolved by SDS-PAGE (7.5% acrylamide gel) and used in Western blotting analysis probed with an anti-VDE polyclonal antibody. ▥ and □, external and internal segments, respectively.

tion of residues at 270–181 with residues at 733–735 was inferred because growth and splicing defects of several conditional mutants carrying mutations at positions 733–735 (*e.g.*, ValLysSer, SerLeuGln and GlyLeuVal) were suppressed by second site mutations of the *VMA1* gene affecting residues 279–281 (Table 2 and Figure 4). Suppression of calcium sensitivity was due to the restoration of the splicing ability (Figure 5B). As expected, some suppressor mutations were located in the VDE region (Asp_{609}Val and Glu_{328}Gly, Figure 4), because the VDE-coding region is thought to contain all elements required for the splicing reaction. Consistent with this idea, a recent random mutagenesis study of the entire VDE sequence identified a His_{362} residue in the VDE-coding region essential for the splicing reaction (KAWASAKI *et al.* 1997). Some mutations that partially affected the splicing reaction were also located at positions around 330 and positions 600–650 (Figure 4) (KAWASAKI *et al.* 1997), suggesting functional involvement of these domains in the reaction. To our surprise, however, mutations of Vma1p (Ile_{279}Asn and Val_{282}Ala) located outside of VDE suppressed the splicing defect caused by mutations at the hydrophobic triplet in the VDE-coding region.

We confirmed functional involvement of the two hydrophobic clusters in the protein splicing reaction using site-specific intragenic suppression. Phenotypic defects of mutations at positions 733–735 (ValLysSer, SerLeuGln and GlyLeuVal) were suppressed by mutations at the positions 279–281 (Table 3), although we could not obtain suppressor mutants from group 4 mutants (GlyLeuTyr, LeuLeuLeu, SerGlyIle and ValMetAsp, data not shown). Western blotting analysis revealed that splicing ability was restored in the sextuple mutants, IleIleTyr to AsnValSer at the positions 279–281 and ValValVal to ValLysSer at the positions 733–735 (Figure

TABLE 3

Amino acid sequence of site-specific suppressor mutants

Sequence at positions		Growth on Ca^{2+}-rich medium at 37°
279–281	733–735	
Ile Ile Tyr	Val Val Val	++ (wild-type)
Ile Ile Tyr	Ser Leu Gln	–
Cys Val Gln	Ser Leu Gln	+
Gly Val His	Ser Leu Gln	+
Leu Val Asn	Ser Leu Gln	+
Phe Val Ser	Ser Leu Gln	+
Ile Gln Tyr	Ser Leu Gln	±
Ile Thr Val	Ser Leu Gln	±
Val Leu Asn	Ser Leu Gln	±
Ile Ile Tyr	Gly Leu Val	–
Asn Ile Ala	Gly Leu Val	+
Cys Ile Lys	Gly Leu Val	+
Ile Val Lys	Gly Leu Val	+
Leu Val Gly	Gly Leu Val	+
Leu Val Ser	Gly Leu Val	+
Met Val Ser	Gly Leu Val	+
Met Val Thr	Gly Leu Val	+
Ser Val Asn	Gly Leu Val	+
Ser Val Asn	Gly Leu Val	+
Val Val Asn	Gly Leu Val	+
Ala Val Ser	Gly Leu Val	±
Gly Val Met	Gly Leu Val	±
Ile Leu Lys	Gly Leu Val	±
Ile Ser Tyr	Gly Leu Val	±
Val Ser Tyr	Gly Leu Val	±
Ile Ile Tyr	Val Lys Ser	–
Asn Val Ser	Val Lys Ser	±
Phe Val Asn	Val Lys Ser	±
Ser Val Asn	Val Lys Ser	±

Amino acid substitutions at positions 733–735 are also listed. In the rightmost column, growth of cells on calcium-rich YPD plates at 37° is indicated as follows: ++, equivalent growth to wild-type; +, weak growth; ±, very weak growth; –, no growth.

6B). Thus, taken together, the results suggest that the two sites play a crucial functional role in the protein splicing reaction.

We also observed that the intragenic synthetic defect occurred when one mutation at the triplet 733–735 was combined with a second mutation at triplet 279–281. A pair of triplet mutations, CysLeuThr at position 733–735 and GluPheLeu at position 279–281 resulted in a severe growth defect and reduced production of the 50-kD VDE protein, although each triplet mutation alone showed little phenotypic defect under the same condition (S. Nogami, Y. Satow, Y. Ohya and Y. Anraku, unpublished data). Moreover, a point mutation of Gly_{283}Val immediately preceding the N-terminal junction inhibited protein splicing and caused accumulation of the 120-kD precursor (S. Nogami, Y. Satow, Y. Ohya and Y. Anraku, unpublished data).

Although our results suggest that the amino acid residues located outside of the VDE-coding region participate in the splicing reaction, their roles supposedly depend on the internal hydrophobic triplet in the C-terminal of VDE. For example, mutations at the positions 279–281 alone had no or little effect on protein splicing (S. Nogami, Y. Satow, Y. Ohya and Y. Anraku, in preparation). Furthermore, the recent folding-dependent *in vitro* experiment showed that VDE bracketed by only six proximal amino acids (including the IleIleTyr sequence that was mutagenized in this study) and four distal amino acids processed autocatalytically (Kawasaki *et al.* 1996). This suggests that most of the regions upstream of VDE are dispensable for the splicing reaction.

Thus, our current findings on the interactions between the hydrophobic residues upstream of the splicing junctions suggest alternative interpretations for previous articles: Cooper *et al.* (1993) reported that the N-terminal half of the Vma1 protozyme they used was dispensable for splicing, however, the artificial N-terminal junction (IleAlaAlaGlnAlaCys instead of the original IleIleTyrValGlyCys) is rather hydrophobic (underlined) in nature. *In vitro* experiments of the archaeal splicing elements embedded in foreign proteins resulted in the production of byproducts reflecting incompleteness of the reaction (Xu *et al.* 1993, 1994). In this case, amino acids at the artificial N-terminal junction (GlyThrArgArgAlaSer instead of the original Lys IleLeuAlaAsnSer) are less hydrophobic (underlined).

Interpretations of genetic interaction between the hydrophobic triplets at the positions 279–281 and 733–735: There are two possible interpretations of genetic interaction between the hydrophobic residues preceding the two splicing junctions: one is that the two hydrophobic clusters interact directly with each other, and the other is that the mutations affect the structure of the folded precursor more globally, rather than locally. Since VDEs with protein splicing mutations at 733–735 retain endonuclease activity (Figure 3), the mutations most likely did not affect the structure of the folded VDE globally, and therefore the latter interpretation seems to be less likely.

We propose a possible model for spatial interactions between residues 279–281 and 733–735 in the protein splicing reaction (Figure 7). This model, which delineates a parallel arrangement of the splicing junctions, is based on and consistent with the following lines of evidence: (1) Residues at positions 279–281 and 733–735 functionally interact. (2) Two cysteines at positions 284 and 738, just after the cleavage sites, must be located in close proximity for progression of the splicing reaction. (3) Residues at both positions 279–281 and 733–735 are predicted to form β-strands by plural algorithms using the SOPMA program (Geourjon and Deléage 1995, data not shown). (4) Preference parameters of Ile and Val to form parallel β-strands are higher than to form antiparallel β-strands (Lifson and Sander 1979), and the presence of Ile and Val pair on adjacent

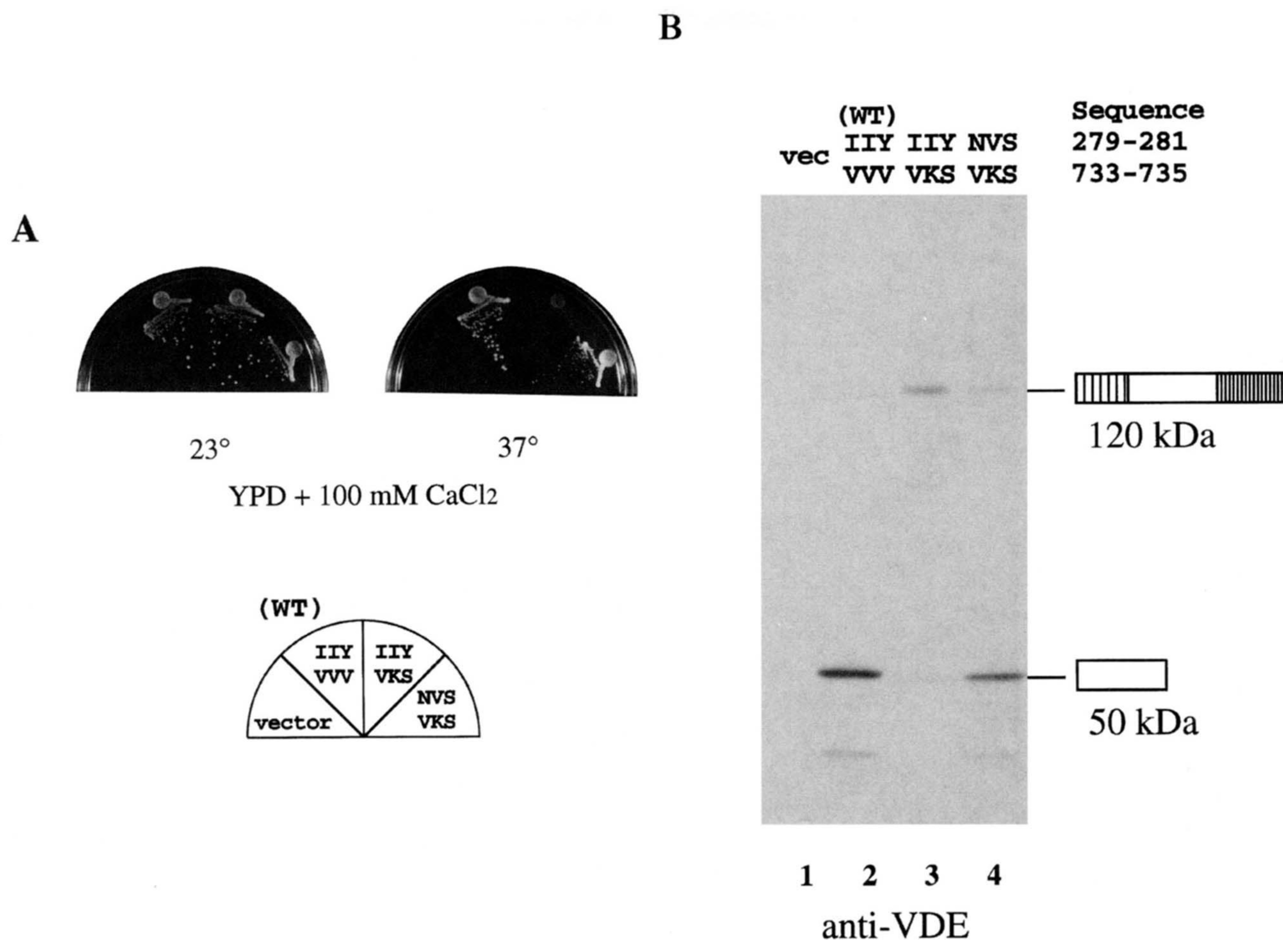


FIGURE 6.—Replacement at the external segments suppresses the splicing defect of the mutations at the internal segment. (A) Growth on YPD + 100 mM $CaCl_2$ plates of strain NY101 transformed with the plasmids pYO314 (vector), pSN101 (WT), and pSN101 carrying the indicated substitutions (one-letter symbol) at the positions 279–281 (top lines) and/or 733–735 (bottom lines) of the *VMA1* gene. Cells were grown at the indicated temperatures on YPD medium supplemented 100 mM $CaCl_2$. (B) Strain NY101 was transformed with the plasmids pYO314 (vec), pSN101 (WT), and pSN101 carrying the indicated substitutions at the positions 279–281 (top lines) and 733–735 (bottom lines) of the *VMA1* gene. Cell extracts prepared from cells cultured in YNBD medium supplemented with casamino acids and buffered at pH 5.0 at 37° were resolved by SDS-PAGE (7.5% acrylamide gel) and used in Western blotting analysis probed with an anti-VDE polyclonal antibody. ▥ and □, external and internal segments, respectively.

β-strands is significantly correlated with parallel β-sheet structures (LIFSON and SANDER 1980). (5) The residues that correspond to the positions 279–281 of the 70-kD Vma1p are considered to form part of a parallel β-sheet in the β-subunit of F1-ATPase from X-ray crystal data (ABRAHAMS *et al.* 1994). (6) The residues of RecA protein that correspond to positions 279–281 of the 70-kD Vma1p are considered to form part of a β-sheet (STORY *et al.* 1992) from X-ray crystal data. In our model, therefore, the upstream regions of the splicing junctions form parallel β-strands (Figure 7B), to assure a proximal contact of the two reaction sites to initiate the splicing reaction. This model does not exclude the possibility that other sites within the VDE-coding region may participate in catalysis of protein splicing (Figure 4, positions 328 and 609, and KAWASAKI *et al.* 1997). Moreover, this model explains the occurrences of hydrophobic residues upstream of the N- and C-terminal splicing junctions of the protozymes (Figure 1) and provides us structural and mechanistic perspectives for the mechanism of protein splicing. The parallel arrangement of the two sites separated by more than 450 amino acid residues correctly aligns the splicing junctions and consistently delineates the steps of the proposed mechanism of the protein splicing pathway (CHONG *et al.* 1996).

Finally, functional involvement of the external segment of VDE in the protein splicing reaction suggests that the VDE element has its own preference for the site to "home" into the host chromosome. The VDE-coding region should insert in a DNA region that encodes a polypeptide segment with the enhancer signal for protein splicing. Without a preferred sequence for insertion, the internal segment might spread without restriction, generating splicing-defective mutations that will sometimes be lethal to the host cells. In this biologi-

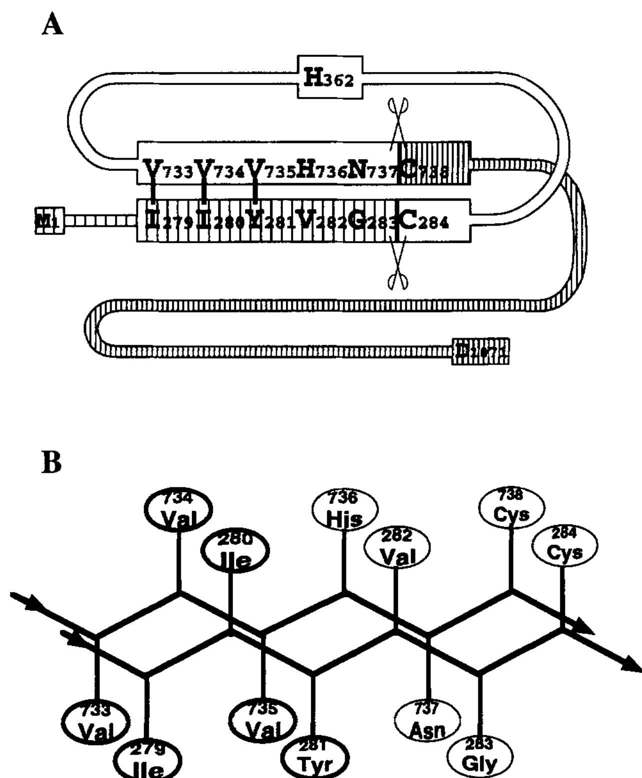


FIGURE 7.—(A) A schematic showing possible spatial interaction between residues 279–281 and 733–735 near the splicing junctions in the Vma1 protozyme. The two hydrophobic triplets are proposed to form parallel β-strands as indicated. Scissors indicate the cleavage sites. Striped and open regions represent the domains of the 70-kD V-ATPase subunit and 50-kD VDE, respectively. His_{362} is another pivotal amino acid in the splicing reaction that was recently located in the middle of VDE (KAWASAKI *et al.* 1997). (B) A structural model for parallel arrangement of two β-strands near the splicing junction. The polypeptide backbones forming the two β-strands are represented as bold lines. Bold ellipses indicate hydrophobic side chains of the two neighboring triplet residues in the interacting β-strands. Side chains of the respective amino acid in the splicing junctions are depicted in ellipses, as positions relative to (top and bottom) the parallel β-strands. Arrows indicate the direction of the main chains of the Vma1 protozyme.

cal context, we are very much interested in knowing how protozymes form an efficient tertiary structure for protein splicing and how symbiotic endonucleases recognize a chromosomal site for noninvasive cleavage and gene homing.

We thank R. HIRATA and M. KAWASAKI for providing plasmids and antibodies and for technical advice and I. SHIMADA and C. NISHIMURA for valuable discussions. This work was supported in part by a grant-in-aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan, and by grants from the Ministry of International Trade and Industry of Japan and the Mitsubishi Foundation, Japan. This work was also supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists to S.N.

LITERATURE CITED

ABRAHAMS, J. P., A. G. W. LESLIE, R. LUTTER and J. E. WALKER, 1994 Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. Nature **370:** 621–628.

ANRAKU, Y. and R. HIRATA, 1994 Protozyme: emerging evidence in nature. J. Biochem. **115:** 175–178.

BULT, C. J., O. WHITE, G. J. OLSEN, L. ZHOU, R. D. FLEISCHMANN *et al.*, 1996 Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science **273:** 1058–1073.

CADWELL, R. C., and G. F. JOYCE, 1992 Randomization of genes by PCR mutagenesis. PCR Methods Appl. **2:** 28–33.

CHONG, S., Y. SHAO, H. PAULUS, J. BENNER, F. B. PERLER *et al.*, 1996 Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an *in vitro* splicing system. J. Biol. Chem. **271:** 22159–22168.

COOPER, A. A., Y.-J. CHEN, M. A. LINDORFER and T. H. STEVENS, 1993 Protein splicing of the yeast *TFP1* intervening protein sequence: a model for self-excision. EMBO J. **12:** 2575–2583.

DAVIS, E. O., P. J. JENNER, P. C. BROOKS, M. J. COLSTON and S. G. SEDGWICK, 1992 Protein splicing in the maturation of *M. tuberculosis* RecA protein: a mechanism for tolerating a novel class of intervening sequence. Cell **71:** 201–210.

DAVIS, E. O., H. S. THANGARAJ, P. C. BROOKS and M. J. COLSTON, 1994 Evidence of selection for protein introns in the RecAs of pathogenic mycobacteria. EMBO J. **13:** 699–703.

FSIHI, H., V. VINCENT and S. T. COLE, 1996 Homing events in the *gyrA* gene of some mycobacteria. Proc. Natl. Acad. Sci. USA **93:** 3410–3415.

GEOURJON, C., and G. DELEAGE, 1995 SOPMA: significant improvements in protein secondary structure prediction by prediction from multiple alignments. Comput. Appl. Biosci. **11:** 681–684.

GIMBLE, F. S., and J. THORNER, 1992 Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae.* Nature **357:** 301–306.

GU, H. H., J. XU, M. GALLAGHER and G. E. DEAN, 1993 Peptide splicing in the vacuolar ATPase subunit A from *Candida tropicalis.* J. Biol. Chem. **268:** 7372–7381.

HIRATA, R., and Y. ANRAKU, 1992 Mutations at the putative junction sites of the yeast Vma1 protein, the catalytic subunit of the vacuolar membrane H^+-ATPase, inhibit its processing by protein splicing. Biochem. Biophys. Res. Commun. **188:** 40–47.

HIRATA, R., Y. OHSUMI, A. NAKANO, H. KAWASAKI, K. SUZUKI *et al.*, 1990 Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H^+-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae.* J. Biol. Chem. **265:** 6726–6733.

HO, S. N., H. D. HUNT, R. M. HORTON, J. K. PULLEN and L. R. PEASE, 1989 Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77:** 51–59.

JONES, J. S., and L. PRAKASH, 1990 Yeast *Saccharomyces cerevisiae* selectable markers in pUC18 polylinkers. Yeast **6:** 363–366.

KAISER, C., S. MICHAELIS and A. MITCHELL, 1994 *Methods in Yeast Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

KANE, P. M., C. T. YAMASHIRO, D. F. WOLCZYK, N. NEFF, M. GOEBL *et al.*, 1990 Protein splicing converts the yeast *TFP1* gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase. Science **250:** 651–657.

KAWASAKI, M., S.-I. MAKINO, H. MATSUZAWA, Y. SATOW, Y. OHYA *et al.*, 1996 Folding-dependent *in vitro* protein splicing of the *Saccharomyces cerevisiae VMA1* protozyme. Biochem. Biophys. Res. Commun. **222:** 827–832.

KAWASAKI, M., S. NOGAMI, Y. SATOW, Y. OHYA and Y. ANRAKU, 1997 Identification of three core regions essential for protein splicing of the Yeast Vma1 protozyme. A random mutagenesis study of the entire *VMA1*-derived sequence. J. Biol. Chem. **272:** 15668–15674.

LAMBOWITZ, A. M. and M. BELFORT, 1993 Introns as mobile genetic elements. Annu. Rev. Biochem. **62:** 587–622.

LIFSON, S., and C. SANDER, 1979 Antiparallel and parallel β strands differ in amino acid residue preferences. Nature **282:** 109–111.

LIFSON, S., and C. SANDER, 1980 Specific recognition in the tertiary structure of β-sheets of proteins. J. Mol. Biol. **139:** 627–639.

MUHLRAD, D., R. HUNTER and R. PARKER, 1992 A rapid method for localized mutagenesis of yeast genes. Yeast **8:** 79–82.

OHYA, Y., N. UMEMOTO, I. TANIDA, A. OHTA, H. IIDA *et al.*, 1991 Calcium-sensitive *cls* mutants of *Saccharomyces cerevisiae* showing a Pet^- phenotype are ascribable to defects of vacuolar membrane H^+-ATPase activity. J. Biol. Chem. **266:** 13971–13977.

PALZKILL, T., and D. BOTSTEIN, 1992 Probing β-lactamase structure and function using random replacement mutagenesis. Proteins **14:** 29–44.

PERLER, F. B., D. G. COMB, W. E. JACK, L. S. MORAN, B. QIANG *et al.*, 1992 Intervening sequences in an Archaea DNA polymerase gene. Proc. Natl. Acad. Sci. USA **89:** 5577–5581.

PERLER, F. B., E. O. DAVIS, G. E. DEAN, F. S. GIMBLE, W. E. JACK *et al.*, 1994 Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. Nucleic Acids Res. **22:** 1125–1127.

PERLER, F. B., G. J. OLSEN and E. ADAM, 1997 Compilation and analysis of intein sequences. Nucleic Acids Res. **25:** 1087–1094.

PIETROKOVSKI, S., 1994 Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. Protein Sci. **3:** 2340–2350.

PIETROKOVSKI, S., 1996 A new intein in cyanobacteria and its significance for the spread of inteins. Trends Genet. **12:** 287–288.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SHIH, C.K., R. WAGNER, S. FEINSTEIN, C. KANIK-ENNULAT and N. NEFF, 1988 A dominant trifluoperazine resistance gene from *Saccharomyces cerevisiae* has homology with F0F1 ATP synthase and confers calcium-sensitive growth. Mol. Cell Biol. **8:** 3094–3103.

SIKORSKI, R. S., and P. HIETER, 1989 A system of shuttle vector and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics **122:** 19–27.

STORY, R. M., I. T. WEBER and T. A. STEITZ, 1992 The structure of the *E. coli recA* protein monomer and polymer. Nature **355:** 318–325.

XU, M.-Q., M. W. SOUTHWORTH, F. B. MERSHA, L. J. HORNSTRA and F. B. PERLER, 1993 In vitro protein splicing of purified precursor and the identification of a branched intermediate. Cell **75:** 1371–1377.

XU, M.-Q., D. G. COMB, H. PAULUS, C. J. NOREN, Y. SHAO *et al.*, 1994 Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. EMBO J. **13:** 5517–22.

Communicating editor: D. BOTSTEIN