The EMBO Journal vol.12 no.6 pp.2575-2583, 1993

Protein splicing of the yeast *TFP1* intervening protein sequence: a model for self-excision

Antony A.Cooper, Yen-Ju Chen, Margaret A.Lindorfer[1] and Tom H.Stevens[2]

Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

[1]Present address: Department of Pathology, University of Virginia Health Science Center, Charlottesville, VA 22908, USA

[2]Corresponding author

Communicated by D.Meyer

Protein splicing is the protein analogue of RNA splicing in which the central portion (spacer) of a protein precursor is excised and the amino- and carboxy-terminal portions of the precursor reconnected. The yeast Tfp1 protein undergoes a rapid protein splicing reaction to yield a spliced 69 kDa polypeptide and an excised 50 kDa spacer protein. We have demonstrated that the 69 kDa species arises by reformation of a bona fide peptide bond. Deletion analyses indicate that only sequences in the central spacer protein of the Tfp1 precursor are critical for the protein splicing reaction. A fusion protein in which only the Tfp1 spacer domain was inserted into an unrelated protein also underwent efficient splicing, demonstrating that all of the information required for protein splicing resides within the spacer domain. Alteration of Tfp1p splice junction residues blocked or kinetically impaired protein splicing. A protein splicing model is presented in which asparagine rearrangement initiates the self-excision of the spacer protein from the Tfp1 precursor. The Tfp1 spacer protein belongs to a new class of intervening sequences that are excised at the protein rather than the RNA level.

Key words: mobile genetic element/protein introns/protein splicing/self-splicing/yeast

Introduction

Protein splicing is one of many processes that modify the informational flow from gene to mature protein. This unusual process is exemplified by the *Saccharomyces cerevisiae TFP1* gene product (Kane *et al.*, 1990). *TFP1* encodes a 119 kDa protein (Tfp1p) that undergoes protein splicing to produce both the 69 kDa catalytic subunit of the vacuolar H^+-ATPase and a 50 kDa spacer protein (Shih *et al.*, 1988; Hirata *et al.*, 1990; Kane *et al.*, 1990; Hendrix, 1991; Hirata and Anraku, 1992). This reaction involves the excision of the intervening spacer protein from the central portion of the 119 kDa precursor protein, and the joining of the N- and C-domains to form the vacuolar H^+-ATPase subunit (Figure 1). Protein splicing of Tfp1p has been shown to occur in *Escherichia coli*, yeast, and when translated *in vitro* (Kane *et al.*, 1990).

Two additional examples of protein splicing have recently been discovered: RecA from *Mycobacterium tuberculosis* (Davis *et al.*, 1991, 1992) and DNA polymerase from the thermophilic Archaebacterium *Thermococcus litoralis* (Hodges *et al.*, 1992; Perler *et al.*, 1992). In each case an intervening amino acid sequence, with homology to the spacer protein of Tfp1p, separates two domains of the mature protein (Shub and Goodrich-Blair, 1992). Both the RecA and DNA polymerase spacer sequences have been shown to be removed post-translationally, and genetic evidence suggests that the majority of the N- and C-domains of the RecA precursor are not required for the splicing reaction (Davis *et al.*, 1992; Hodges *et al.*, 1992).

A seemingly less related post-translational polypeptide rearrangement occurs in the maturation of the plant lectin concanavalin A (Carrington *et al.*, 1985). The process involves the cleavage and formation of peptide bonds, yet it is different from the above cases of protein splicing. For concanavalin A, the reaction results in reversing the order of the precursor's N- and C-domains rather than excising a large intervening protein sequence (Bowles *et al.*, 1986; Bowles and Pappin, 1988).

The excised Tfp1p spacer protein has recently been identified as a highly specific DNA endonuclease that cleaves a site in a *TFP1* allele that is created by the exact deletion of the spacer DNA (*TFP1-spacer*Δ allele; Bremer *et al.*, 1992; Gimble and Thorner, 1992). *In vitro* studies with the purified spacer protein [designated VDE by Gimble and Thorner (1992)] demonstrated that cleavage occurred within the *TFP1-spacer*Δ DNA at the N/C domain junction, but the spacer protein did not cleave the wild-type *TFP1* DNA. Cleavage of the *TFP1-spacer*Δ gene by the spacer protein in a *TFP1/TFP1-spacer*Δ heterozygote was shown to initiate gene conversion that converted an allele that lacked the intervening sequence (*TFP1-spacer*Δ) into one that contained it (*TFP1*). These observations demonstrated that the spacer protein is capable of mediating the movement of its encoding DNA sequence, thereby indicating that the *TFP1* intervening sequence is genetically mobile.

We investigated the mechanism of protein splicing and report here that the spacer protein can splice from a completely unrelated insertional context. Mutational analysis of the residues at the Tfp1p splice junctions reveals that certain residues are critically important for the protein splicing reaction. To account for our findings, we propose a protein splicing model involving self-excision of the spacer domain.

Results

Protein splicing joins the N- and C-domains via a peptide bond

Previous experiments have indicated that the *TFP1*-encoded 119 kDa precursor (Tfp1p) undergoes protein splicing to produce the 69 kDa vacuolar H^+-ATPase subunit (Kane

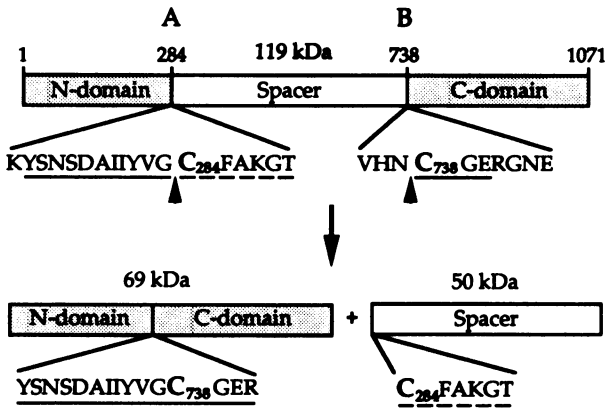

Fig. 1. Protein splicing of yeast Tfp1p. The schematic diagram shows the 119 kDa Tfp1p precursor protein undergoing a splicing reaction at the protein level to produce both the 50 kDa spacer protein and the 69 kDa subunit of the vacuolar H^+-ATPase. Shown are the residues at splice junctions A and B, the protein sequence for the peptide that spans the splice junction of the 69 kDa subunit (solid underlined) and the amino-terminal protein sequence of the spliced spacer protein (dashed underlined). The amino acid sequence of the peptide spanning the spliced junction and the spacer protein amino terminus were determined by protein sequencing. The splice junction cysteines are numbered relative to the initiating methionine codon of *TFP1*. Arrowheads indicate proposed cleavage points in Tfp1p.

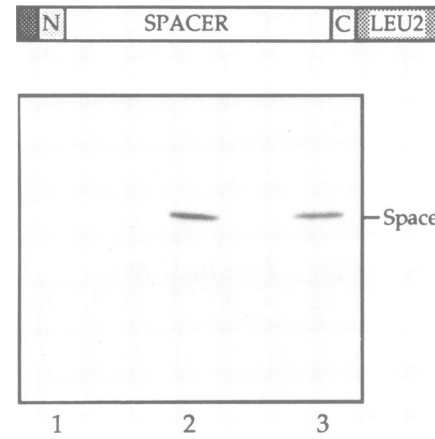

Fig. 2. Large deletions of the N- and C-domains do not inhibit protein splicing. The schematic diagram shows the predicted protein encoded by pAAC100: the 12 residues comprising the c-*myc* epitope, the distal 28 residues of the N-domain (N), the complete spacer domain (SPACER), the proximal 13 residues of the C-domain (C) and 150 residues encoded by the *LEU2* gene (LEU2). The strain SEY6211a-*tfp1*Δ was transformed with the following plasmids: lane 1, pRS316 (centromere containing vector with no insert; Sikorski and Heiter, 1989); lane 2, pPK26 (pRS316 containing *TFP1*): lane 3, pAAC100 (pRS316 containing the c-*myc*-N-spacer-C-Leu2p gene fusion). Cells were grown to mid-log phase in liquid YEP media containing raffinose and galactose (2% final concentration) for 6 h prior to harvesting. Cell extracts were prepared as described, resolved by SDS–PAGE and used in immunoblots probed with affinity-purified anti-spacer antibodies.

et al., 1990). In such a reaction, the 50 kDa spacer protein is excised from the central portion of the precursor, while the N- and C-domains are joined to create the vacuolar H^+-ATPase subunit (Figure 1). Indirect evidence had suggested that the joining of the N- and C-domains is via the formation of a peptide bond (Kane *et al.*, 1990). To test this prediction, tryptic peptides from the native 69 kDa vacuolar H^+-ATPase subunit were separated by HPLC to identify the peptide spanning the junction. Several peptides that eluted near the position calculated for the junction peptide were sequenced and all agreed with the predicted amino acid sequence of regions of the N- and C-domains of the 69 kDa polypeptide. The relevant peptide was identified and is shown in Figure 1 (solid underlined), as are the amino acid sequences at the two splice junctions of Tfp1p. Edman degradation of the bond joining the N- and C-domains demonstrates that the domains are linked via a peptide bond.

Only one cysteine residue was detected in the sequenced junction peptide from the 69 kDa vacuolar H^+-ATPase subunit, yet a cysteine residue is present at each splice junction in the Tfp1p precursor (C284 and C738; Figure 1). Mechanistically, it is important to determine which of the two cysteine residues remains in the 69 kDa polypeptide as it allows one to assign the peptide bonds that are broken in the precursor and reformed in the spliced product. To identify the position of the cysteine in the 50 kDa spacer protein, this polypeptide was purified and subjected to amino-terminal sequencing. Cysteine, and the subsequent sequence shown in Figure 1 (dashed underlined), was found at the amino terminus of the spacer protein, thereby defining the precise breakage points (arrowheads, Figure 1) and assigning C738 to the spliced 69 kDa polypeptide.

Removal of the N- and C-domains does not affect protein splicing

To test the role of the Tfp1p N- and C-domains in protein splicing, deletions were constructed in either or both domains. Large deletions in either the N- or C-domains did not prevent protein splicing (data not shown). A chimeric construct (pAAC100) was produced that combined the separate deletions of the N- and C-domains. Given that the N- and C-domains were now very small, the c-*myc* epitope and a portion of yeast *LEU2* gene were added to tag the regions flanking the spacer by epitope or mass addition. In addition to these unrelated sequences, the chimeric construct encoded the complete Tfp1p spacer flanked by the distal 28 residues of the N-domain and the proximal 13 residues of the C-domain (Figure 2). The fusion protein was expressed in a strain disrupted at the *TFP1* locus (*tfp1*Δ) and Western blot analysis was performed on protein extracts with antibodies directed against the spacer protein. If protein splicing occurred at the predicted junctions, then the spacer protein would be excised as a 50 kDa protein. If splicing failed to occur, a 70 kDa protein is predicted that could be detected with both anti-spacer protein and anti-c-*myc* antibodies. No proteins were identified from the strain containing the vector alone (Figure 2, lane 1), whereas the strain expressing the fusion protein produced a 50 kDa protein, which was detected with anti-spacer protein antibodies and co-migrated with authentic spacer protein (Figure 2, lanes 2 and 3). The excision of the 50 kDa spacer protein from such a construct predicts that a c-*myc*-tagged 20 kDa protein should result from the splicing reaction. However, the anti-c-*myc* monoclonal antibody failed to detect any protein resulting from expression of pAAC100 (data not shown), suggesting that the expected c-*myc*-tagged 20 kDa protein was unstable in yeast.

The spacer domain can splice from a new context

Truncations of the majority of both the N- and C-domains did not affect splicing, and raised the possibility that the

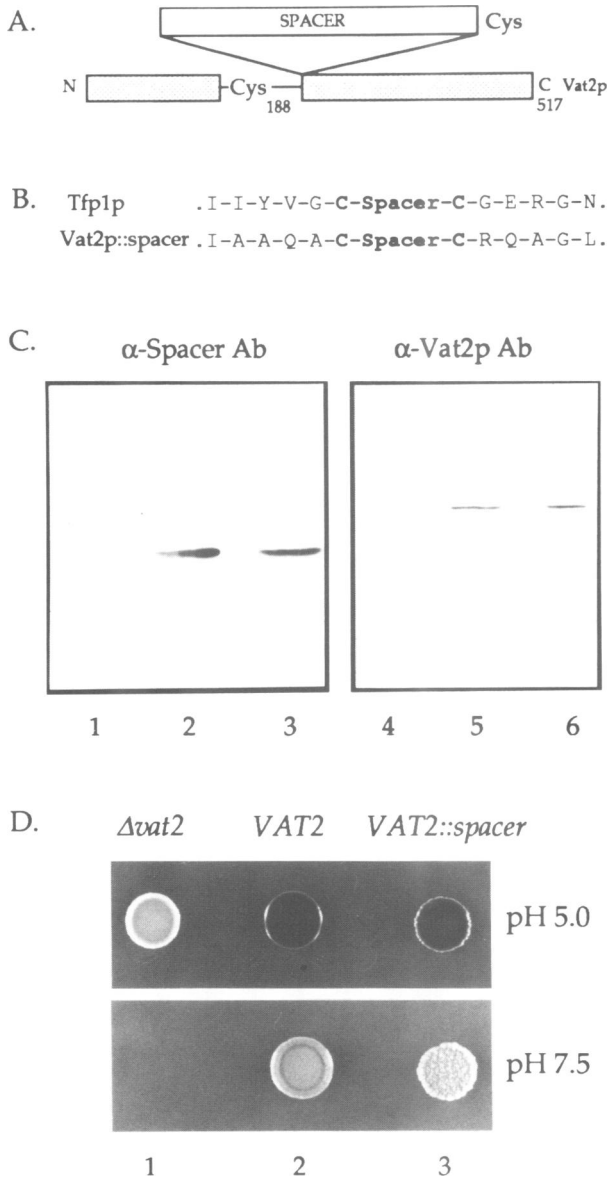

Fig. 3. The spacer protein is sufficient for protein splicing. (**A**) The schematic diagram shows the Vat2p::spacer fusion protein construct. (**B**) Shown are the amino acid residues flanking the spacer protein in the context of either Tfp1p or Vat2p::spacer. (**C**) The strain SEY6211a-*tfp1*Δ contained the following plasmids: lane 1, pRS316 (vector); lane 2, pPK26 (pRS316 containing *TFP1*); lane 3, pAAC108 (pRS316 containing *VAT2*::*spacer*). Cell extracts were prepared as described, resolved by SDS−PAGE and used in immunoblots probed with affinity-purified anti-spacer antibodies. The strain SEY6211a-*vat2*Δ contained the following plasmids: lane 4, pRS316; lane 5, pAAC101 (pRS316 containing *VAT2*); lane 6, pAAC108. The resulting immunoblots were probed with an anti-Vat2p monoclonal antibody. (**D**) The SEY6211a-*vat2*Δ strain contained the following plasmids: column 1, pRS316; column 2, pAAC101 (*VAT2* in pRS316); column 3, pAAC108 (*VAT2*::*spacer* in pRS316). The cells were grown at 30°C on low-adenine synthetic selective media buffered to either pH 5.0 or 7.5.

functional domain required for Tfp1p to undergo protein splicing might be contained completely within the spacer protein itself. To test this hypothesis, a gene fusion was constructed in which the coding region for the spacer protein was precisely inserted into the open reading frame of the yeast *VAT2* gene adjacent to a cysteine codon (Cys_{188}; Figure 3A). The *VAT2* gene encodes the 60 kDa subunit

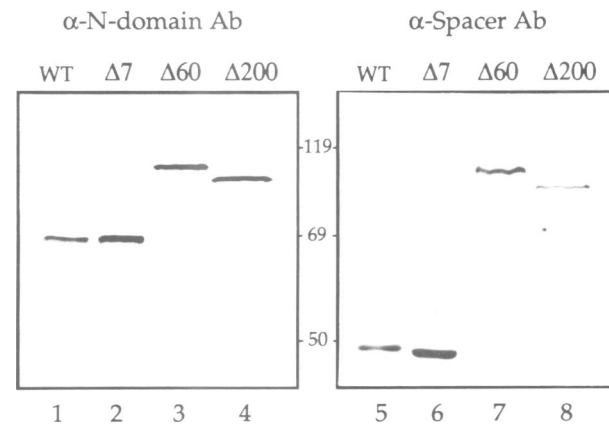

Fig. 4. Deletions within the spacer domain prevent splicing. Strain SEY6211a-*tfp1*Δ was transformed with the following plasmids: pPK26 (*TFP1*, lanes 1 and 5), pAAC50 (Δ7, lanes 2 and 6), pAAC51 (Δ60, lanes 3 and 7) and pAAC52 (Δ200, lanes 4 and 8). Cell extracts were prepared as described, resolved by SDS−PAGE and used in immunoblots probed with either anti-N-domain monoclonal antibody (lanes 1−4) or affinity-purified anti-spacer antibodies (lanes 5−8). The spacer codons deleted in these constructs are described in Materials and methods.

(Vat2p) of the vacuolar H^+-ATPase (Nelson *et al.*, 1989; Yamashiro *et al.*, 1990) and shares no significant sequence similarity with the *TFP1*-encoded 69 kDa subunit. As seen in Figure 3B, the amino acid residues flanking the spacer protein in its native *TFP1* context and in the *VAT2*::*spacer* construction are unrelated. The *VAT2*::*spacer* gene fusion carried on a centromere-based plasmid (pAAC108; Figure 3A) was transformed into both *vat2*Δ and *tfp1*Δ strains, and Western blot analysis performed on protein extracts from these strains to identify the proteins produced. Anti-spacer protein antibodies identified a 50 kDa protein from the *tfp1*Δ strain carrying plasmid pAAC108 that co-migrated with authentic spacer protein (Figure 3C, lanes 2 and 3). Monoclonal antibodies directed against Vat2p detected a 60 kDa protein from the *vat2*Δ strain expressing pAAC108 that co-migrated with wild-type Vat2p (Figure 3C, lanes 5 and 6). These results demonstrate that the spacer protein is capable of splicing from a completely different insertional context.

In order to determine if the 60 kDa Vat2p encoded by the *VAT2*::*spacer* allele was functional, we tested the ability of the *VAT2*::*spacer* gene to complement a *vat2*Δ mutation. Yeast mutants lacking a subunit of the vacuolar H^+-ATPase (*vat2*Δ, *tfp1*Δ, etc.) are sensitive to the pH of the growth medium; they can grow in medium buffered to pH 5.0, but not pH 7.5 (Yamashiro *et al.*, 1990). Figure 3D shows that the *vat2*Δ strain expressing either wild-type Vat2p or the Vat2p::spacer fusion protein was capable of growing at both pH 5.0 and 7.5, whereas the strain carrying the vector alone was incapable of growth at pH 7.5. In addition to a pH sensitivity, *vat2* mutants, in an *ade2* genetic background, are white as opposed to the usual red color (Foury, 1990). The *vat2*Δ strain shown in Figure 3D carries an *ade2* mutation and the cells expressing plasmid-borne Vat2p or the Vat2p::spacer fusion protein produce red colonies on pH 5.0 plates, whereas the cells carrying the vector alone produce white colonies. We therefore conclude that excision of the spacer domain from the Vat2p::spacer fusion protein results in a spliced Vat2p that is functionally and bio-

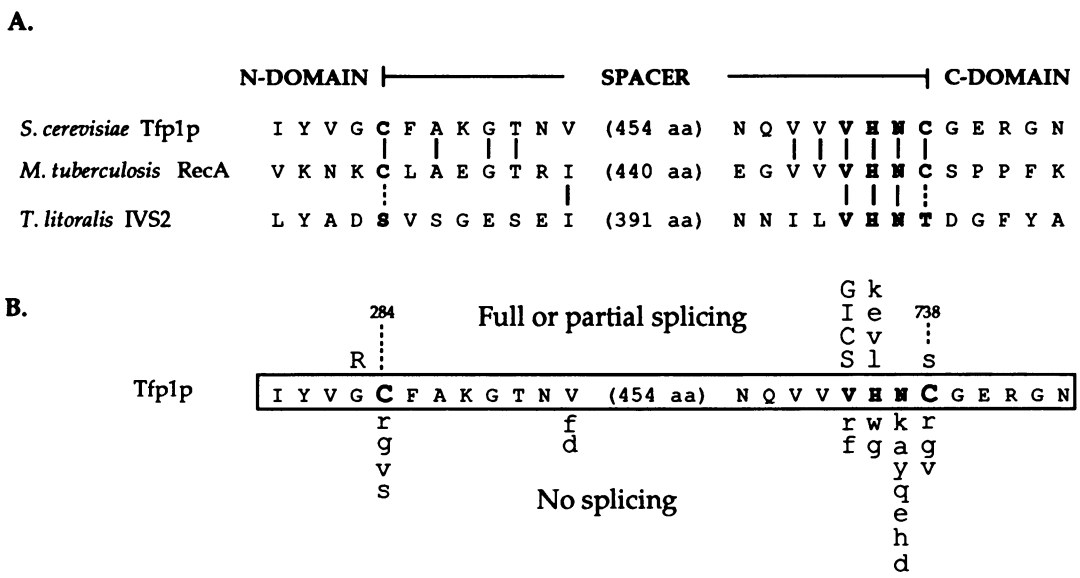

Fig. 5. Homology and substitution of amino acid residues at the splice junctions. (**A**) Shown are the splice junction residues of the three known examples of protein splicing. Vertical bars show conserved residues, while dotted lines indicate semi-conserved residues. The number of amino acid residues contained within the respective spacer domains is indicated. (**B**) The schematic diagram shows the amino acid residues present at the splice junctions of Tfp1p (boxed). Substitution of these residues: (i) allows >90% splicing (residues represented in upper case above the box); (ii) inhibits splicing so that ≤20% of Tfp1p is spliced (residues represented in lower case above the box); or (iii) blocks splicing completely (residues shown below the box). Also shown are the amino acid residue numbers at the splice junctions of Tfp1p.

chemically indistinguishable from Vat2p encoded by the non-interrupted *VAT2* gene.

Deletions within the spacer domain prevent protein splicing

The excision of the spacer domain from the Vat2p::spacer protein suggested that the information required for protein splicing is contained within the spacer domain and implied that mutations within the spacer domain would prevent splicing. A mutational analysis was performed by constructing in-frame deletions at a site in the middle of the 454 amino acid (aa) spacer region (at Tfp1p codon 513). Small in-frame insertions or deletions (Δ7 residues) at this site did not affect splicing. As expected, the Δ7 construct produced a correctly sized 69 kDa polypeptide, whereas the resulting spacer protein was smaller in size (Figure 4, lanes 2 and 6). Larger in-frame deletions (Δ60 and Δ200 residues) prevented splicing and produced correspondingly truncated, but stable, precursors (Figure 4, lanes 3, 4, 7 and 8). Deletions at other positions within the spacer domain also prevented splicing (data not shown). These non-splicing forms containing in-frame spacer deletions (Δ60, Δ200) did not splice when co-expressed with wild-type *TFP1* (data not shown).

Junction residues play a pivotal role in splicing

Recently, two additional examples of protein splicing have been discovered: DNA polymerase from the thermophilic Archaebacterium *T.litoralis* (Hodges *et al.*, 1992; Perler *et al.*, 1992) and RecA from *M.tuberculosis* (Davis *et al.*, 1991, 1992). In each case, the central spacer domains are 42–50 kDa in size and share sequence homology (21–23% amino acid identity between any two), particularly at splice junction B where the motif VHNC/T is found (Figure 5A).

A number of amino acid substitutions were created at the Tfp1p splice junctions, and the effects of several such mutations (C284S, V735G, H736G, N737Q, C738S) are shown in the Western blot probed with an anti-N-domain

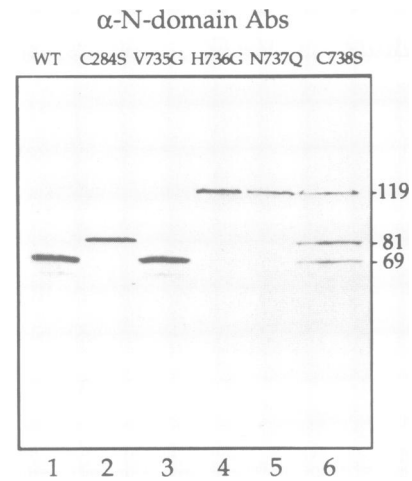

Fig. 6. The substitution of Tfp1p junction residues affects protein splicing. Strain SEY6211a-*tfp1*Δ was transformed with the plasmid pPK26 encoding the following substitutions at the Tfp1p splice junctions: WT (lane 1), C284S (lane 2), V735G (lane 3), H736G (lane 4), N737G (lane 5) and C738S (lane 6). Cell extracts were prepared as described, resolved by SDS–PAGE and used in immunoblots probed with anti-N-domain monoclonal antibody.

monoclonal antibody (Figure 6). Splicing was highly sensitive to the substitution of particular residues (C284, N737), while other residues (V735, H736, C738) could be altered to create a spectrum of phenotypes ranging from slight inhibition to a complete blockage of splicing. Figure 5B summarizes the substitutions tested and indicates whether they blocked splicing. Altered forms of Tfp1p with amino acid substitutions represented in upper case exhibit >90% of Tfp1p splicing to mature products in the steady state (Figure 5B). The substitutions represented in lower case above the Tfp1p sequence allowed limited splicing (≤20%), whereas the alterations below the sequence prevented the formation of any spliced product. A species of intermediate

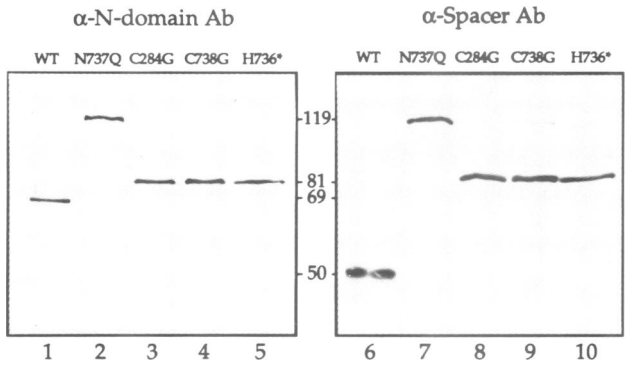

Fig. 7. Substitution of certain splice junction residues produces cleaved but not spliced products. Strain SEY6211a-*tfp1*Δ was transformed with pPK26 (*TFP1*, lanes 1 and 6) or with pPK26 containing the following mutations: N737Q (lanes 2 and 7), C284G (lanes 3 and 8), C738G (lanes 4 and 9) and H736* (a stop codon substituted for H737, lanes 5 and 10). Cell extracts were prepared as described, resolved by SDS−PAGE and used in immunoblots probed with either anti-N-domain monoclonal antibody (lanes 1−5) or affinity-purified anti-spacer antibodies (lanes 6−10).

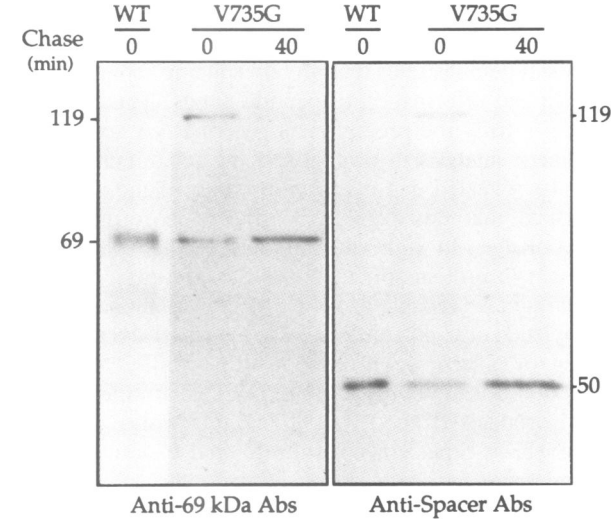

Fig. 8. Protein splicing of mutant Tfp1p can occur post-translationally. Strain SEY6211a (WT) or strain SEY6211a-*tfp1*Δ transformed with pPK26 containing the mutation V735G were radiolabeled for 5 min, whereupon half the culture was harvested (0 chase), while the remainder of the culture was incubated in the presence of excess unlabeled methionine and cysteine for an additional 40 min. The cells were lyzed as described and the denatured proteins were immunoprecipitated with either affinity-purified anti-vacuolar H^+-ATPase 69 kDa sununit or affinity-purified anti-spacer protein antibodies. The precipitated samples were analyzed by SDS−PAGE and fluorography.

size (81 kDa) was detected for some of the mutant forms of Tfp1p (Figure 6, lanes 2 and 6) and the implications of this observation were investigated further (see below).

Protein splicing may initiate at splice junction B

N737 plays a critical role in protein splicing, as shown by the finding that substitution of any amino acid resulted in non-spliced products (Figure 5B; Figure 7, lane 2). Certain substitutions of other junction residues produced proteins of a size intermediate between that of the precursor and spliced products. Alteration of either of the splice junction cysteine residues to glycine produced a species that corresponded to a precursor that had undergone a cleavage event at splice junction B (Figure 7, lanes 3 and 4) to yield the N-domain-spacer domain (N-Sp) lacking the C-domain. The identification of the protein as an N-Sp species was based on several lines of evidence: it corresponded to the predicted size of 81 kDa (737 aa), and was detected by antibodies directed against the spacer protein and N-domain (Figure 7), but not by anti-C-domain antibodies. Instead, anti-C-domain antibodies detected a separate 37 kDa protein, which corresponds to the size predicted for the C-domain (data not shown). In addition, the 81 kDa species co-migrated in SDS−PAGE with a mutant Tfp1p truncated at residue 736 (Tfp1p-H736*; Figure 7, lanes 5 and 10). The discovery of the N-Sp species suggested that cleavage at junction B may initiate the splicing reaction. Consistent with this hypotheses is the observation that the truncation mutant Tfp1p-H736*, which is missing junction B, shows no cleavage at junction A, while the addition of 15 residues to its C-terminus (2 spacer and 13 C-domain residues; pAAC100; Figure 2) restored junction B and allowed splicing to occur.

Protein splicing is post-translational

The rapid rate of protein splicing has precluded a kinetic demonstration of a precursor−product relationship for wild-type Tfp1p. However, Western blot analysis of cells expressing a mutant allele of *TFP1*, V735G, detected a small amount of precursor in addition to the spliced products (Figure 6). A pulse−chase analysis demonstrated that the mutation slowed the splicing reaction. As expected, after a 5 min labeling period, there was no detectable unspliced wild-type Tfp1 precursor (Figure 8, lanes 1 and 4). In contrast, at the end of the 5 min labeling period, Tfp1p-V735G was present both as the unspliced 119 kDa precursor and as the spliced 69 kDa and 50 kDa polypeptides (Figure 8, lanes 2 and 5). During a subsequent 40 min chase period, the 119 kDa Tfp1p precursor was quantitatively converted to the two spliced protein products with a half-time of ~15 min. These data demonstrate that the splicing reaction can proceed post-translationally.

Discussion

In this paper, we make four major advances in our knowledge concerning protein splicing. First, we demonstrate for the first time that a peptide bond is indeed reformed during the protein splicing reaction. Second, we have elucidated the precise site of peptide bond cleavage and formation in the precursor and spliced products. Third, we have demonstrated that the Tfp1p spacer protein undergoes protein splicing when placed in a new protein context, indicating that all of the information required for splicing is contained within the spacer protein, and we have accounted for both of the splicing reaction products. Fourth, we have demonstrated the importance of residues at the Tfp1p splice junctions and, in particular, observed a strict requirement for asparagine at the second splice junction. In addition, we present a protein splicing model involving self-excision that invokes an extraordinary role for asparagine in the initiation of this novel post-translational reaction.

The first point was established by amino acid sequencing of the tryptic peptide spanning the spliced junction of the 69 kDa H^+-ATPase subunit. Edman degradation of this

peptide confirmed that a bona fide peptide bond connects the spliced N- and C-domains within the 69 kDa polypeptide. This result demonstrates that protein splicing represents the cleavage and reformation of peptide bonds. Thus, protein splicing is truly the protein analogue of RNA splicing.

A kinetic analysis of protein splicing for a mutant form of Tfp1p revealed a precursor−product relationship and established that this peptide bond can be formed post-translationally. In light of the dispensable role of the C-domain in protein splicing, it is possible that wild-type Tfp1p might be capable of splicing co-translationally prior to completion of C-domain synthesis. Co-translational protein splicing of Tfp1p would explain the inability to follow conversion of the wild-type 119 kDa Tfp1p protein to the spliced products (Kane *et al.*, 1990; this work).

Removal or replacement of the N- and C-domains does not affect the correct excision of the intervening spacer domain from the Tfp1p precursor. In fact, the spacer protein was capable of splicing from a new insertional context (Vat2p), which bore no sequence similarity to the native insertional site within the 69 kDa polypeptide. This result indicates that any potential *trans*-acting splicing machinery responsible for the excision event must be capable of acting independently of the spacer domain context. Such a candidate could involve a protease with a recognition site containing an invariant asparagine. Alternatively, the spacer domain may mediate its own excision in the absence of additional protein factors. Consistent with a self-excision model are the observations that protein splicing occurred when *TFP1* was expressed in *E.coli*, the yeast cytoplasm and several *in vitro* translation systems (Kane *et al.*, 1990; Ryan *et al.*, 1992). In addition, Tfp1p also underwent protein splicing when it was targeted and translocated into the yeast endoplasmic reticulum (Y.-J.Chen and T.H.Stevens, unpublished observation), which is a highly oxidizing environment quite different from that of the cytoplasm (Hwang *et al.*, 1992). Further support for a general self-splicing model (not limited to Tfp1p) are the findings that *M.tuberculosis* RecA and *T.litoralis* DNA polyermase undergo protein splicing in the relevant native organism as well as when the genes are expressed in *E.coli* or in an insect cell line (Davis *et al.*, 1992, Hodges *et al.*, 1992). These results suggest that either the potential *trans*-acting protein(s) involved in protein splicing is extremely conserved across the biological kingdoms and present in different subcellular compartments or processing occurs by a self-splicing mechanism. Consistent with a self-splicing model is that the splice junctions are not the sole determinants of protein splicing, since deletions in the middle of the Tfp1p, RecA or DNA polymerase spacer region produce stable unspliced precursors containing wild-type splice junction sequences (Figure 4, this work; Davis *et al.*, 1992; Hodges *et al.*, 1992). Finally, the inability to isolate mutations in extragenic loci that block splicing of Tfp1p in yeast is also consistent with a self-splicing model (K.J.Hill and T.H.Stevens, unpublished data).

Model for protein splicing

The Tfp1p precursor contains a cysteine residue at each splice junction, yet only one of these is present in the mature 69 kDa polypeptide (Figure 1). A simple model predicts that the remaining cysteine junction residue is present at either the amino or carboxy terminus of the spacer protein. Purification and sequencing of the spliced spacer protein demonstrated that cysteine exists at the amino terminus of this protein. These results indicate that at splice junction B of Tfp1p (Figure 1), where splicing is thought to be initiated, a peptide bond is broken between the conserved asparagine (N737) and cysteine (C738) residues. In addition, a peptide bond must be broken between G283 and C284 to position a cysteine residue at the amino terminus of the spacer protein, and a peptide bond would be formed between G283 and C738 to create the 69 kDa vacuolar H^+-ATPase subunit.

The strict requirement for asparagine at the proposed initiating splice junction of Tfp1p suggests a model based on the ability of asparagine residues to cause peptide bond cleavage (Geiger and Clarke, 1987; Clarke *et al.*, 1992). Under physiological conditions, the asparagine β-amide nitrogen is capable of attacking the peptide bond carbonyl, resulting in cleavage of the peptide bond and formation of a C-terminal succinimide ring (Figure 9A). Thus, it is possible for an asparagine residue to spontaneously form an intramolecular succinimide ring, the result of which is the breakage of the peptide bond carboxy terminal to the asparagine residue. Such a cleavage reaction has been shown to occur in both peptides and proteins (Voorter *et al.*, 1988; Violand *et al.*, 1990; Clarke *et al.*, 1992). The succinimide model for the initiation of Tfp1p splicing is consistent with the finding that no amino acid would substitute for N737. It is possible that the structure of the spacer within the precursor provides an optimal fixed alignment of the asparagine side-chain nitrogen and peptide carbonyl so as to promote extremely rapid succinimide ring formation. In almost all situations, it would be deleterious for a protein to attain a conformation compatible with succinimide-mediated peptide bond cleavage except, as in the case of Tfp1p, where rapid cleavage is required.

The model we propose involves the asparagine residue (N737) forming a succinimide ring in the manner indicated, resulting in the spontaneous breakage of the peptide bond linking the asparagine residue to the neighboring cysteine (Figure 9). In this model, the folded structure of the spacer domain within Tfp1p brings the two splice junctions together (as with self-splicing introns), so that in a subsequent, or possibly concerted step, the cysteine residue at the amino terminus of the C-domain (C738) is available to undergo a transpeptidation reaction with the G283-C284 peptide bond at splice junction A (Figure 9B). Alternatively, the structure of the N- and C-domains of Tfp1p may be responsible for bringing the splice junctions into close proximity. However, since removal of all but 28 residues of the N-domain and 13 residues of the C-domain of Tfp1p does not interfere with efficient splicing (Figure 2), the protein splicing reaction cannot depend on unique N- and C-domain structures.

A partial test of this succinimide-mediated excision model involves the resolution of the predicted succinimide at the C-terminus of the spacer protein following splicing. Hydrolysis of this structure occurs rapidly and is predicted to result in a mixture of aspartic acid amide and asparagine residues (Clarke *et al.*, 1992). We are currently attempting to determine if aspartic acid amide is present at the C-terminus of the spliced 50 kDa spacer protein, thereby demonstrating that a succinimide intermediate is involved in the protein splicing reaction.

It is too early to propose a specific role for the junction

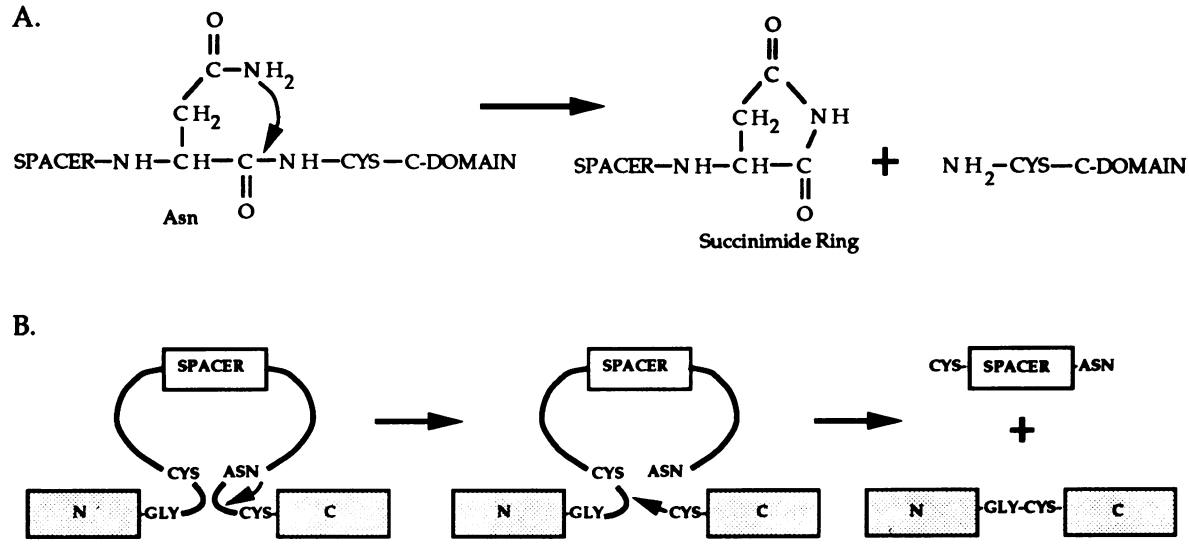

Fig. 9. Model for protein splicing of Tfp1p. (**A**) Shown schematically is the nucleophilic attack of the Asn β nitrogen on the peptide bond carbonyl to create a succinimide ring and cleavage of the peptide bond linking N737 and C738. (**B**) The schematic diagram indicates, in the context of Tfp1p, that the spacer domain structure brings the two splice junction Cys residues close in space. Following the succinimide-mediated cleavage of the peptide bond at splice junction B, the liberated C738 attacks the G283-C284 peptide bond in a transpeptidation reaction to release the spacer protein and create the G283-C738 peptide bond. Although an intermediate step is presented for clarity, it is quite possible that both the succinimide-mediated cleavage and transpeptidation are part of a concerted reaction.

cysteine residues; however, their role is clearly important since alteration of either cysteine to glycine resulted in cleavage between the spacer and C-domain, but no splicing of the N- and C-domains of Tfp1p. It is conceivable that a disulfide bond forms between the cysteines at the splice junctions or, alternatively, that the two sulfhydryls share a bound divalent metal ion. The presence of serine and threonine at the junctions of the *T.litoralis* I-*Tli*I spacer suggests that residues other than cysteine can function at the junctions of protein splicing elements. However, the substitutions of serine and threonine at the splice junctions may reflect the elevated growth temperature (85–90°C) of the Archaebacterium and/or an altered redox state in the cytoplasm of this organism.

An alternative mechanism for protein splicing has been proposed that involves the motif His-Asn-Cys (Thr/Ser) at junction B, resembling the 'catalytic triad' found in cysteine and serine proteases (Hodges *et al.*, 1992). This model proposes that the histidine residue activates the cysteine or serine residue, which proceeds to attack the relevant peptide bond at junction B. The substitutions presented here (Figure 5) demonstrate that, at least in the case of Tfp1p, the histidine residue is not essential since substitution with either Lys, Glu, Val or Leu allows partial to near wild-type levels of splicing to occur.

The 50 kDa spacer protein has previously been shown to be a double-stranded DNA endonuclease (Gimble and Thorner, 1992). Apart from this post-splicing role, the spacer domain within the Tfp1p precursor is likely to perform additional functions. The model we propose involves the spacer assuming a conformation within the precursor that is optimal for rapid succinimide formation of splice junction B. In addition, the spacer structure is proposed to bring the two splice junctions together to allow a transpeptidation step to proceed. This model does not exclude the possibility that amino acids in the spacer region, besides those at the splice junctions, may participate in catalyzing the protein splicing reaction.

A novel class of mobile genetic elements

A direct analogy exists between the proposed self-excision of the Tfp1p spacer protein and the excision of self-splicing RNA introns (Belfort, 1990; Cech, 1990; Jacquier, 1990). Intervening sequences present in DNA can be inherently excised either by an RNA splicing mechanism or post-translationally by protein splicing, thereby allowing such elements to remain phenotypically silent.

A further similarity shared between Group I introns and the examples of protein splicing is genetic mobility (Shub and Goodrich-Blair, 1992). The Tfp1p spacer protein has recently been identified as a highly specific DNA endonuclease that can mediate the insertion of the spacer encoding sequence into alleles of *TFP1* that lack the spacer sequence (Gimble and Thorner, 1992). The transfer of the intervening sequence DNA by gene conversion is very similar to a process that occurs in a family of Group I introns that encode endonucleases within the intron, where it has been demonstrated that these endonucleases endow the introns with mobility (Lambowitz, 1989; Perlman and Butow, 1989; Belfort, 1990). In a process referred to as 'intron homing', the DNA encoding such an intron and its internally encoded endonuclease inserts at high frequency into a recipient sister allele that lacks the intervening sequence.

The spacer proteins from Tfp1p and the *T.litoralis* DNA polymerase (I-*Tli*I) have been found to possess a DNA endonuclease activity (Gimble and Thorner, 1992; Perler *et al.*, 1992). Neither the endonuclease activity nor the mobility function has been tested for the intervening sequence of RecA from *M.tuberculosis*, although it also contains homology to the yeast HO endonuclease. The spacer protein of *S.cerevisiae TFP1* therefore highlights the existence of

a new family of intervening sequences that undergo excision by protein splicing and have the capacity to be genetically mobile.

Materials and methods

Strains, growth conditions and materials

SEY6211a-*tfp1*Δ was created by disrupting the *TFP1* locus of SEY6211a (*MAT***a** *ura3-52 leu2-3,112 his3*-Δ200 ade2-101 trp1-Δ901 suc2-Δ*9*) with the plasmid pPK8 cut with *Xba*I (Kane *et al.*, 1990). SEY6211a-*vat2*Δ was created by disrupting the *VAT2* locus of SEY6211a with plasmid pCY40 cut with *Hin*dIII (Yamashiro *et al.*, 1990). BJ3505 (*MAT***a** *pep4*::*HIS3 prb1-Δ1.6R lys2-208 trp1-Δ101 ura3-52 gal2 can1*) was obtained from the Yeast Genetic Stock Center (Berkeley).

Unless otherwise noted, cells were grown in YEPD medium or synthetic dextrose (SD) medium with the appropriate supplements (Sherman *et al.*, 1982). Low-adenine medium consists of SD with the appropriate supplements, except that the adenine concentration has been lowered to 3 mg/l.

Mutagenesis and plasmids

The sequence encoding the c-*myc* epitope (MEQKLISEEDLF) was inserted downstream of the *GAL1* promoter previously inserted into pRS316 (Sikorski and Hieter, 1989). A 1.5 kb fragment encoding residues 256–752 of Tfp1p was inserted downstream of the c-*myc* sequence. Finally, a 1.8 kb *Eco*RI–*Sac*I fragment of the yeast *LEU2* gene (containing codons 214–364) was ligated to the 3′ termini of the *TFP1* fragment to produce plasmid pAAC100.

Oligonucleotide-directed mutagenesis was performed (Kunkel *et al.*, 1987) on *VAT2* to introduce an *Sph*I site at bp +558, changing the protein sequence from Ile_{187}-Cys to Ala_{187}-Cys; this mutant allele fully complemented the *vat2*Δ mutation. *Sph*I sites were also introduced at the sequence encoding both splice junctions of *TFP1*, changing Gly_{283}Cys to Ala_{283}Cys and Asn_{737}Cys to Ala_{737}Cys. The 1.3 kb *Sph*I fragment containing the spacer encoding sequence was inserted into the introduced *Sph*I site within *VAT2*, creating pAAC107. Mutagenesis was performed to change the Ala_{737}Cys sequence in the *VAT2*::*spacer* allele to the wild-type sequence of Asn-Cys, thereby creating construct pAAC108.

The in-frame spacer deletions were produced by the introduction of a *Nru*I site at bp +1539 (numbered with respect to the initiation codon of *TFP1*) in the plasmid pPK26 using site-directed mutagenesis with the oligonucleotide 5′-CAAAGAAGTATCTCGCGAATCAACCGAAAA-3′. ExoIII deletion was performed on the *Nru*I-cleaved plasmid (Erase-a-Base, Promega) and the resulting library of deletions transformed into SEY6211a-*tfp1*Δ. Colonies were chosen and screened by Western blot for the production of spacer protein with altered molecular mass. Mutations of interest were identified by sequencing, or in the case of Δ60, estimated by fine restriction fragment mapping. The deletion Δ7 is missing residues 514–520, while Δ200 is lacking residues 513–712.

Amino acid substitutions at the splice junctions were achieved via oligonucleotide-directed mutagenesis performed on pPK26 ssDNA using the following oligonucleotides: 5′-TATGTCGGGNNCTTTGCCAAG-3′ (C284X); 5′-CAGGTTGTCHNCCATAATTGCG-3′ (V735X); 5′-GTTGCGTCDNVAATTGCCGGAG-3′ (H736X); T5′-TGTCGTCCATNANTGCGGAGAAA-3′ (N737X); 5′-TGTCGTCCATCAGTGCGGAGAAAG-3′ (N737Q); 5′-C-GTCCATAATNNCGGAGAAAG-3′ (C738X). The resulting libraries were transformed into *E.coli* from which clones were sequenced by double-stranded sequencing using Sequenase (USB). Plasmids containing the desired mutations were transformed into SEY6211a-*tfp1*Δ to determine if the mutant Tfp1p was capable of (i) complementing the *tfp1*Δ mutation and (ii) to what degree the mutant Tfp1p could splice.

Other mutations at splice junction A were obtained by oligonucleotide-directed mutagenesis performed on pPK26R (pPK26 with the *TFP1* insert in the opposite orientation) with the oligonucleotide 5′-TAAAACATTGGTACCCTTGGCAAAGCACCCGACATAGATA-3′ doped to 1.35% at each position. Following mutagenesis, the *E.coli* clones were pooled and the extracted DNA transformed into SEY6211a-*tfp1*Δ and plated onto low-adenine SD medium. White colonies were selected and Western blot analyses were performed on cell extracts to identify non-spliced products. Plasmid DNA was rescued (Hoffman and Winston, 1987) from interesting clones and sequenced as described above.

Antibodies, Western blots and immunoprecipitations

Anti-N domain (8B1) monoclonal, anti-Vat2p monoclonal (13D11) and anti-spacer polyclonal antibodies have been described previously (Kane *et al.*, 1990). Polyclonal antibodies directed against the 69 kDa subunit were produced by inserting the 2.3 kb *Stu*I–*Sal*I fragment of the *TFP1-spacer*Δ allele into pEXP3, expressing this in *E.coli* and injecting the resulting fusion protein into rabbits as described previously (Raymond *et al.*, 1990). The crude serum was affinity purified against the same fusion protein conjugated to cyanogen bromide-activated Sepharose (Sigma). Cell extracts, SDS–PAGE and Western blots were performed as described by Yamashiro *et al.* (1990).

Cell labeling was performed by growing cultures in supplemented minimal media lacking methionine to an OD_{600} of 1.0, then pulse labeled by the addition of 200 μCi/ml Express label (DuPont) for the desired time. Chase conditions were achieved by the addition of methionine and cysteine to a final concentration of 50 μg/ml. At each time point, azide was added to 10 mM and the cells spheroplasted as described by Stevens *et al.* (1986), except that spheroplasting was performed with oxalyticase (200 μg/ml; Enzogenetics) at 30°C for 5 min prior to boiling in 1% SDS. Immunoprecipitations were performed as described previously (Roberts *et al.*, 1992).

Amino acid sequencing

The 69 kDa subunit of the vacuolar H^+-ATPase was isolated, digested with trypsin and the tryptic peptides separated by HPLC (Aebersold *et al.*, 1987; Ho *et al.*, 1993). Preliminary sequencing data showed that the splice peptide co-eluted with several other peptides. The experiment was repeated, putative splice peptide fractions collected, pooled and reduced in volume to 50 μl; 100 μl of 6 M GuHCl, 0.25 M Tris (pH 8), 1 mM EDTA were added. The solution was reduced by addition of 2 μl of 10% β-mercaptoethanol and incubated for 2 h under argon at room temperature. The mixture was then alkylated by the addition of 2 μl of neat vinylpyridine and incubated for 2 h under argon at room temperature, protected from light (Friedman *et al.*, 1970). The reaction was stopped by acidification with 15 μl of 10% trifluoroacetic acid (TFA) and the mixture rechromatographed as before. Several peaks were resolved after the large mercaptoethanol peak eluted. Fractions were collected manually and sequenced with an Applied Biosystems Model 475 sequencing cycles provided by the manufacturer.

For amino-terminal sequencing, the spacer protein was overexpressed in strain BJ3505 carrying *TFP1* on the 2 μm based vector pSEY8 and partially purified from a cell lysate. Purification of the spacer protein was achieved by ammonium sulfate precipitation and column chromatography as described by F.S.Gimble and J.Thorner (personal communication). The material was then precipitated by acetone (4 vol at −20°C), resuspended in 250 mM Tris (pH 8.6), 0.66% SDS and 200 mM β-mercaptoethanol, and placed in the dark at room temperature under argon for 2 h. 4-Vinyl-pyridine (Aldrich) was added to a final concentration of 500 mM and the incubation continued for an additional 2 h. The alkylating agent was removed by centrifuging the protein solution through a spin column (Isolab) containing G25 Sephadex (Sigma), and the protein further purified by preparative SDS–PAGE and electroelution into 40 mM CAPS (pH 9.8). The eluted material was bound to a PVDF membrane by means of a Pro-Spin column (Applied Biosystems) and subjected to protein sequencing.

Acknowledgements

We thank Steven Clarke for alerting us to the instability of Asn in proteins and Rick Dahlquist for stimulating discussions. We are very grateful to Fred Gimble for providing the purification procedure for the spacer protein prior to publication and to Deb McMillen for sequencing the amino terminus of the spacer protein. We also thank Rick Dahlquist, Diane Hawley and members of the Stevens and Sprague labs for critical reading of the manuscript. A.A.C. was supported by a fellowship from the American Heart Association (Oregon Affiliate). This work was supported by a grant from the American Cancer Society (VM-33) and an American Cancer Society Faculty Research Award to T.H.S.

References

Aebersold,R.H., Leavitt,J., Saavedra,R.A., Hood,L.E. and Kent,S.B.H. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6970–6974.

Belfort,M. (1990) *Annu. Rev. Genet.*, **24**, 363–385.

Bowles,D.J. and Pappin,D.J. (1988) *Trends Biochem.*, **13**, 60–64.

Bowles,D.J., Marcus,S.E., Pappin,D.J.C., Findlay,J.B.C., Eliopoulos,E., Maycox,P.R. and Burgess,J. (1986) *J. Cell Biol.*, **102**, 1284–1297.

Bremer,M.C.D., Gimble,F.S., Thorner,J. and Smith,C.L. (1992) *Nucleic Acids Res.*, **20**, 5484.

Carrington,D.M., Auffret,A. and Hanke,D.E. (1985) *Nature*, **313**, 64–66.

Cech,R.T. (1990) *Annu. Rev. Biochem.*, **59**, 543–568.
Clarke,S., Stephenson,R.C. and Lowenson,J.D. (1992) In Manning,M.C. and Berchan,R. (eds), *Pharmaceutical Biotechnology*. Plenum Press, New York.
Davis,E.O., Sedgwick,S.G. and Colston,M.J. (1991) *J. Bacteriol.*, **173**, 5653–5662.
Davis,E.O., Jenner,P.J., Brooks,P.C., Colston,M.J. and Sedgwick,S.G. (1992) *Cell*, **71**, 201–210.
Foury,F. (1990) *J. Biol. Chem.*, **265**, 18554–18560.
Friedman,M., Krull,L.H. and Covins,J.F. (1970) *J. Biol. Chem.*, **245**, 3868–3871.
Geiger,T. and Clarke,S. (1987) *J. Biol. Chem.*, **262**, 785–794.
Gimble,F.S. and Thorner,J. (1992) *Nature*, **357**, 301–306.
Hendrix,R.W. (1991) *Curr. Biol.*, **1**, 71–73.
Hirata,R. and Anraku,Y. (1992) *Biochem. Biophys. Res. Commun.*, **188**, 40–47.
Hirata,R., Ohsumi,Y., Nakano,A., Kawasaki,H., Suzuki,K. and Anraku,Y. (1990) *J. Biol. Chem.*, **265**, 6726–6733.
Ho,M.N., Hill,K.J., Lindorfer,M.A. and Stevens,T.H. (1993) *J. Biol. Chem.*, **268**, 221–227.
Hodges,R.A., Perler,F.B., Noren,C.J. and Jack,W.E. (1992) *Nucleic Acids Res.*, **20**, 6153–6157.
Hoffman,C.S. and Winston,F. (1987) *Gene*, **57**, 267–272.
Hwang,C., Sinskey,A.J. and Lodish,H.F. (1992) *Science*, **257**, 1496–1502.
Jacquier,A. (1990) *Trends Biochem. Sci.*, **15**, 351–354.
Kane,P.M., Yamashiro,C.T., Wolczyk,D.F., Goebl,M., Neff,N. and Stevens,T.H. (1990) *Science*, **250**, 651–657.
Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Cell*, **56**, 323–326.
Nelson,H., Mandiyan,S. and Nelson,N. (1989) *J. Biol. Chem.*, **264**, 1775–1778.
Perler,F.B. *et al.* (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5577–5581.
Perlman,P.S. and Butow,R.A. (1989) *Science*, **246**, 1106–1109.
Raymond,C.K., O'Hara,P.J., Eichinger,G., Rothman,J.H. and Stevens,T.H. (1990) *J. Cell Biol.*, **111**, 877–892.
Roberts,C.J., Nothwehr,S.F. and Stevens,T.H. (1992) *J. Cell Biol.*, **119**, 69–83.
Ryan,C., Stevens,T.H. and Schlesinger,M.J. (1992) *Protein Sci.*, **1**, 980–985.
Sherman,F., Fink,G.R. and Hicks,J.B. (1982) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Shih,C.-K., Wagner,R., Feinstein,S., Kanik-Ennulat,C. and Neff,N. (1988) *Mol. Cell. Biol.*, **8**, 3094–3103.
Shub,D.A. and Goodrich-Blair,H. (1992) *Cell*, **71**, 183–186.
Sikorski,R.S. and Hieter,P. (1989) *Genetics*, **122**, 19–27.
Stevens,T.H., Rothman,J.H., Payne,G.S. and Schekman,R. (1986) *J. Cell Biol.*, **102**, 1551–1557.
Violand,B.N., Schlittler,M.R., Toren,P.C. and Siegel,N.R. (1990) *J. Protein Chem.*, **9**, 109–117.
Voorter,C.E.M., de Haard-Hoekman,W.A., van den Oetelaar,P.J.M., Bloemendal,H. and de Jong,W.W. (1988) *J. Biol. Chem.*, **263**, 19020–19023.
Yamashiro,C.T., Kane,P.M., Wolczyk,D.F., Preston,R.A. and Stevens,T.H. (1990) *Mol. Cell. Biol.*, **10**, 3737–3749.

Received on January 19, 1993; revised on March 8, 1993