Review

Reflections on protein splicing: structures, functions and mechanisms

By Yasuhiro ANRAKU*1,2,† and Yoshinori SATOW*3

(Communicated by Takao SEKIYA, M.J.A.)

Abstract: Twenty years ago, evidence that one gene produces two enzymes via protein splicing emerged from structural and expression studies of the VMA1 gene in *Saccharomyces cerevisiae*. VMA1 consists of a single open reading frame and contains two independent genetic information for Vma1p (a catalytic 70-kDa subunit of the vacuolar H\textsuperscript{+}-ATPase) and VDE (a 50-kDa DNA endonuclease) as an in-frame spliced insert in the gene. Protein splicing is a posttranslational cellular process, in which an intervening polypeptide termed as the VMA1 intein is self-catalytically excised out from a nascent 120-kDa VMA1 precursor and two flanking polypeptides of the N- and C-exteins are ligated to produce the mature Vma1p. Subsequent studies have demonstrated that protein splicing is not unique to the VMA1 precursor and there are many operons in nature, which implement genetic information editing at protein level. To elucidate its structure-directed chemical mechanisms, a series of biochemical and crystal structural studies has been carried out with the use of various VMA1 recombinants. This article summarizes a VDE-mediated self-catalytic mechanism for protein splicing that is triggered and terminated solely via thiazolidine intermediates with tetrahedral configurations formed within the splicing sites where proton ingress and egress are driven by balanced protonation and deprotonation.

Keywords: protein splicing, VMA1, VDE, thiazolidine intermediate, intein, extein

Introduction

In 1990, the Anraku Laboratory discovered evidence that the *Saccharomyces cerevisiae* VMA1 gene\(^1\) expresses the nascent 120-kDa translational product of VMA1 from which the 50-kDa VDE endonuclease (VMA1-derived endonuclease\(^2\)) or the VMA1 intein is excised out and the N- and C-terminal exteins thus formed are ligated to produce the mature 70-kDa catalytic subunit of the vacuolar H\textsuperscript{+}-ATPase (see Figs. 1 and 2). This posttranslational splicing is rapid and found to occur both *in vivo* and *in vitro* systems.\(^3\) Since the first discovery of the VMA1 intein, more than 100 putative inteins have been identified in eubacteria, archaea, and eukaea species.\(^4\)

The comprehensive studies conducted by the Anraku group have shown that the reaction is folding-dependent and auto-catalytic, and takes place within the single molecule.\(^5\)–\(^7\) The mutagenesis study also indicated that the N-extein residues play unique and important roles in protein splicing.\(^8\) The Perler, Xu, and Paulus laboratories have demonstrated that the reaction involves several chemical steps including N\textsuperscript{-}S acyl shift, transesterification, succinimide formation, and S\textsuperscript{-}N acyl shift.\(^9\)–\(^11\) A large amount of biochemical and physicochemical information has been accumulated and eventually one recognizes that the splicing reaction is controlled in a rapid, accurate, and solely structure-dependent manner and also exquisitely coordinated so as to yield the correctly spliced protein as the primary product. Side-reactions such as cleavage at either of the two splicing junctions seem to occur only when inteins are inserted into foreign contexts or made to function under nonphysiological conditions.\(^9\)–\(^12\)–\(^15\) Contem-
porary reviews on these subjects are available in References.\textsuperscript{16–18})

In \textsuperscript{1999}, the Anraku group adopted finally a crucial strategy for dissecting chemical mechanisms by which protein splicing regulates its complicated stepwise processes faithfully without any formation of side products. Crystal structural studies were thus implemented after designing splicing-inactive and slowly spliceable VDE recombinant precursors in soluble forms to address a precise chemical mechanism for protein splicing.\textsuperscript{19–21)}

1. The \textit{VMA1} gene

Nucleotide sequencing of \textit{VMA1} in a \textit{BamHI-XhoI} fragment from the yeast hybridizing clone RH151 (nucleotide residues at –585 to +3497) revealed that the gene consists of a single open reading frame capable of encoding a polypeptide of 1,071 amino-acid residues (ref. 1; see also Fig. 2). The most upstream ATG (nucleotide residues at 1–3) among four possible candidates was determined to be the initiation codon based on the predicted N-terminal amino-acid sequence (residues 1–40). In the 5’-flanking region of the sequence, the TATA-like sequence at nucleotide position –79 and the consensus sequences for the yeast transcription initiation site at nucleotide positions –24 and –20 are located. The whole sequence stretches over 3,213 base-pairs without the yeast mRNA splicing consensus sequences. In the 3’-flanking region, a set of sequences homologous to the yeast transcription termination signal is found.

The homology search indicated that the N- and C-terminal regions of the nascent precursor (residues at Met1–Gly283 and Cys738–Asp1,071, respectively) showed high identity to those of the catalytic subunits of carrot\textsuperscript{22}) and \textit{Neurospora crassa}\textsuperscript{23}) vacuolar membrane H\textsuperscript{+}-ATPases. The in-frame insert with 454 amino-acid residues at Cys284–Asn737 was determined to become a VDE portion after splicing\textsuperscript{9,14}).

Based on these pieces of structural information, the

\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{N-Extein (30 kDa)} & \textbf{Intein (50 kDa VDE)} & \textbf{C-Extein (40 kDa)} & \textbf{Splicing} & \textbf{Crystal} \\
\hline
\textit{VMA1} & M—NYSNSDAIIYCGFAKG—\textbf{T}E—VVVHNCGERNEMAEVL—D & ++ & – \\
\hline
\textbf{XC-VDE-His} & MIYYCVGFAKG—\textbf{T}E—VVVHNCGERHIVIIVHGG & ++ & – \\
\hline
\textbf{MIYYVG-VDE-CGER-Cerm} & MIYYCVGFAKG—\textbf{T}E—VVVHNCGERNEMAEVL—D & ++ & – \\
\hline
\textit{N}_{	ext{new}}-\textit{MIYYVG-VDE-CGER} & M—\textbf{NYSNSDAIIYCGFAKG}—\textbf{T}E—VVVHNCGER & ++ & – \\
\hline
\textbf{XA-VDE} & MIYYCVGFAKG—\textbf{T}E—VVVHNCGER & – & + \\
\hline
\textbf{X10SAS} & M—\textbf{NYSNSDAIIYCGFAKG}—\textbf{T}E—VVVHAGGERNEMAE & – & – \\
\hline
\textbf{X10SSS} & M—\textbf{NYSNSDAIIYCGFAKG}—\textbf{T}E—VVVHAGGERNEMAE & – & + \\
\hline
\textbf{X10SNS} & M—\textbf{NYSNSDAIIYCGFAKG}—\textbf{T}E—VVVHAGGERNEMAE & + & – \\
\hline
\textbf{X10SNC} & M—\textbf{NYSNSDAIIYCGFAKG}—\textbf{T}E—VVVHAGGERNEMAE & ++ & – \\
\hline
\textbf{X10CNS} & M—\textbf{NYSNSDAIIYCGFAKG}—\textbf{T}E—VVVHAGGERNEMAE & ++ & – \\
\hline
\end{tabular}

Fig. 2. Sequences of the \textit{VMA1} protein and recombinants. The N- and C-terminal sequences of the recombinants are aligned with the \textit{VMA1} sequence of which the residues from Cys284 to Asn737 correspond to the VDE intein. Residues subjected to the mutation studies are indicated with bold characters. The N- and C-terminal side polypeptides are indicated with \textit{N} and \textit{C}, respectively. The replacements in the X10-series VDE recombinants are underlined. Splicing activity and crystallization for each recombinant are indicated with a corresponding sign (+ or –) in the columns of “Splicing” and “Crystal”.

Fig. 1. \textit{VMA1} gene and products of the protein splicing.
VMA1 gene is an ordinary operon containing information for two independent enzymes without any redundancy or shortage of genetic codons and transcribed to yield a single species of mRNA.1)

2. Essential residues for protein splicing

The first mutagenic study24) and alignment search for the VMA1 intein family25) suggested that the excision takes place at the peptide bonds between the splicing junctions of Gly283-Cys284 and Asp737-Cys738 (see also Fig. 2). This prediction was soon confirmed by the chemical determination of the N- and C-terminal amino-acid residues of VDE.9),14)

Since the splicing reaction is catalyzed by the in-frame insert, VDE, Kawasaki et al. carried out a systematic search for crucial amino acids involved in the whole process of the reaction.6) A random mutagenesis strategy was used to identify such residues throughout the entire region of VDE and its neighboring elements and seven defective-mutants with single amino-acid replacements were mapped systematically in the three core regions: Gly283Val and Cys284Tyr at the N-terminal splicing junction and another two (Asn737Ser and Asn737Lys) are at the C-terminal splicing junction,6) in agreement with the previous results.14),24)

The other three mutations are new and assigned at amino-acid positions His362, Val403, and Ser639 at the third core region. The His362 is conserved in every predicted splicing precursor,25) and the His362Leu mutation gives a severer defect in protein splicing though a recombinant VDE constructed from the His362Leu mutant, expressed in E. coli and purified shows normal DNA endonuclease activity.6) This finding indicates that active sites for the protein splicing and endonuclease activities reside apart in different regions. The His362 residue contributes to the first cleavage at the N-terminal splicing junction6) whereas the Val403 and Ser639 residues do not participate in an excision-ligation reaction but they have structural roles in regulation of the splicing activity.21)

3. Biochemical features of protein splicing

Kawasaki et al. established a folding-dependent in vitro splicing assay with the use of various VDE recombinant precursors and proved that the splicing reaction catalyses the joining of N-extein Gly283 and C-extein Cys738.9) In addition, with the use of purified XC-VDE-His (see Fig. 2) as a substrate, the splicing reaction proceeds accurately in the neutral pH even in the presence of strong protease inhibitors.7)

To examine a possibility whether the reaction occurs within a single molecule or needs two molecules for splicing, two different species of VDE propeptides (MIIYVG-VDE-CGER — C-terminus and N-terminus — MIIYVG-VDE-CGER; see Fig. 2) were mixed in an equal amount for splicing assay. No swapping of the exteins was taken place as no intact 120-kDa VMA1 precursor was detected, supporting that the reaction is an intra-molecular process.7) All the results indicate that the splicing reaction involves bond rearrangements rather than bond cleavage followed by resynthesis. In other words, a new chemical principle that has never been uncovered in nature is adopted for the world of protein splicing.5),7),14)

4. Schematic diagram for a mechanism of protein splicing reaction

We have proposed a schematic diagram for the VDE-mediated protein splicing reaction (refs. 21, 26; see Fig. 3). In the initial N → S acyl shift, the thiazolidine intermediate at the Gly283-Cys284 junction is formed and then resolved into the thioester intermediate formed at Cys284. In the next step, the branched intermediate13) is formed by the transesterification that involves a nucleophilic attack on the thioester carbon atom by the side chain of conserved Cys738. In the third step, the intervening intein region from Cys284 to Asn737 is excised out by the peptide-bond cleavage that is triggered by a succinimide formation at conserved Asn737. The transient thioester linkage between the exteins undergoes the final S → N acyl rearrangement via the second thiazolidine intermediate.

5. Design of VDE recombinants for structure-directed splicing

The structure of the VDE protein expressed in E. coli27) referred to as generic VDE hereafter, revealed a bilobed domain structure: one domain bears residues requisite for the endonuclease activity, and the other those for the protein splicing. In the protein-splicing domain, the conserved N- and C-terminal residues Cys284, His736 and Asn737, as well as His362 that is essential for the splicing reaction,6) are assembled in a location suitable for taking part in the reaction.

The unspliced XC-VDE recombinant bearing six N-extein and four C-extein residues was expressed
largely as insoluble inclusion bodies.\(^5\) Hence, Mizutani and Satow designed the soluble X10-series VDE recombinants for crystallographic studies, by adopting the extended segments of 11 N-extein and 10 C-extein residues from the \(VMA1\) sequence (ref. 20; see also Fig. 2). In order to reduce reaction speed in this series, mutations of Cys284Ser and His362Asn are introduced in the X10SNC recombinant; His362Asn and Cys738Ser in the X10CNS; Cys284Ser, His362Asn, and Cys738Ser in the X10SNS; and Cys284Ser, His362Asn, Asn737Ser, and Cys738Ser in the X10SSS.

Purified X10-series samples were subjected to protein splicing assay.\(^20\) The X10SSS product gave a single 53-kDa band in the SDS-PAGE analysis and showed a MALDI-TOF mass of 53,010, in good agreement with the predicted mass of 53,078 for the intact X10SSS. Therefore, the X10SSS retains its N- and C-extein propeptides, and is a splicing-inactive precursor.

The X10CNS showed that the product exists mainly as a 50-kDa fraction accompanied by a minor fraction of 53 kDa in the SDS-PAGE.\(^20\) The purified 53-kDa product was incubated for 12 hrs at 4°C and ultrafiltrated with 10-kDa cutoff. The molecular mass of the filtrate fraction was determined to be 2,100 with the mass spectrometry, in exact agreement with that predicted for the spliced peptide (SNSDAILYVGSGERGNEMAE). The N-terminal amino-acid sequence of the filtrate was determined to be SNSDA, indicating that the protein splicing certainly but slowly occurred in the X10CNS. The X10SNC product was also observed as 50-kDa and 53-kDa bands in
the SDS-PAGE. The X10SNS product was purified as a single-band protein of 53 kDa, whereas the 50-kDa band appeared after the crystallization.

6. Three-dimensional structural analyses

Crystals of the X10SSS and X10SNS recombinants were prepared at 20°C within a few days after the purification, and grew to mature sizes within a week. The X10SSS crystals were dissolved into a solution, and the resultant solution gave again the 53-kDa band in the SDS-PAGE. The X10SNS crystals were obtained after the crystallization for 3 and 7 days at 20°C, and the dissolved solutions gave the 50-kDa band. The amount of the 50-kDa band increased with time, indicating that the splicing reaction proceeded slowly even in the crystalline lattice isomorphous to the X10SSS crystal, and then underwent the splicing process, releasing the spliced peptide.

The three-dimensional structure of the X10SSS crystal flash-cooled to 110 K in diffraction experiments was determined at 2.1 Å resolution (Protein Data Bank accession code of 1JVA, ref. 20). The X10SNS structure from the data obtained at 31 hrs after the crystallization setup was similarly obtained at 2.9 Å resolution (accession code 1UM2, ref. 21).

The structures of two molecules A and B in the asymmetric unit in the space group P1 are nearly identical as shown in Fig. 4, consisting of the domain I (residues 463–699), the DNA recognition subdomain of the domain II (residues 369–437), and the splicing subdomain of the domain II (residues 284–368, 438–462, and 700–737). The residues at the C-terminal-side ends of helices α4 (486–501) and α7 (595–609) bear the endonuclease activity, and supposedly retain the domain I structure requisite for the endonuclease activity. Between the X10SSS, X10SNS, and generic VDE27 structures, the splicing junctions give marginal differences. These indicate that the overall structures of the splicing junctions are virtually unalterable, even in the course of the splicing reaction, by the N- and C-extein extension residues (see Figs. 4, 5 and 6). The splicing junctions of the X10SSS and X10SNS are superposed for comparison in Fig. 7.

6.1. Precursor structure of the splicing junctions. The protein-splicing site is situated at the bottom of the β-sheet region of the domain II (Fig. 4). In the X10SSS precursor, the N-extein residues of Val282 and Gly283 are in a β-strand conformation, and located in the vicinity of Asn362 and Ser738 as expected (Fig. 5). The Val282 O atom is hydrogen-bonded to the Asn359 side-chain. Atomic contacts are observed among hydrophobic residues Ile280, Val282, Pro324, and Tyr714. The scissile peptide bond between Gly283 and Ser284 is in an ordinary trans conformation, and its Ser284 N atom is hydrogen-bonded also to the Asn359 side-chain. This trans conformation was also observed in the structure of the XA-VDE19 that bears an Ala284 residue and extra N-extein residues MIIVGY adopted from the VMA1 sequence, and also in the VMA29 VDE structure28 that has an Ala284 residue, extra N-terminal residues MKAEEGKLEG from the maltose binding protein, and extra C-extein residues CGER from the VMA1 sequence.

The main-chain conformation of Ser738 in the C-extein is stabilized by the hydrogen bonds between its N and Ile717 O atoms and between its O and Thr718 Oγ1 atoms (Fig. 5). Residues 739–741 are in the β-strand conformation, forming a short anti-
parallel β-sheet along with the N-extein residues 281–283. The formation of this β-sheet brings the residues of the N- and C-terminal junctions in close proximity; bringing the Gly283 carbonyl O atom into contact with the Ser738 side-chain Oγ atom. The side chains of Glu740 and Arg741 are in electrostatic interaction with residues Lys322 and Glu443, respectively. These Glu443 and Arg741 side-chains are clustering together with charged residues of Asp399, Arg401, Arg445, and Asp446.

The mutagenesis study indicated that the N-extein residues of Ile279, Ile280, and Tyr281 possibly interact with the residues Val733, Val734, and Val735 that precede the C-terminal junction. The strands Ile279–Tyr281 and Val733–Val735 are not directly involved in atomic contact, but brought into a cluster.

Fig. 5. Stereo drawing of the splicing site of the X10SSS precursor. The N- and C-extein residues are drawn in red and yellow, respectively, and their linkages to the intein in green are indicated by blue bonds. Hydrogen bonds involved with the splicing junctions are indicated by dotted light blue lines, with their donor-acceptor distances in Å.

Fig. 6. Stereo view of the splicing site of the X10SNS spliced product molecules. The electron-density map contoured at the 3σ level by omitting 281–283 and 738–741 residues is superposed on the extein residues of molecule B. The N- and C-extein residues are ligated by the Gly283-Ser738 peptide bond.
together with hydrophobic residues Pro324, Ile717, and Phe727. This hydrophobic cluster contributes to the conformational construction of these strands.

6.2. Ligated structure of the splicing junctions. The ligated extein segments in the splicing junctions (residues 282–285, 359–362, and 736–739) are clearly delineated in the X10SNS structure as shown in Fig. 6. Asn737 of the excised intein was reported to turn up in an approximately equimolar amount of asparagine and succinimide forms. The Asn737 structure is interpretable either as an asparagine or as a succinimide, but an asparagine is better fitted into the density (Fig. 6). The mutated His362Asn has its side-chain conformation almost identical to that of the His362 side-chain of the generic VDE.

The Gly283, Ser738, and Gly739 region of the ligated segment is in an atypical turn conformation (Fig. 6). This turn structure involves a hydrogen bond between the Gly283 N and Gly739 O atoms as in the X10SSS structure. The Ser738 main-chain is stabilized also by the hydrogen bond to the Thr718 Oγ1. Residues 739–741 form a short antiparallel β-sheet along with the N-extein residues 282 and 283 (Fig. 6). This β-sheet is originated from the Gly283-Ser738 linkage formed by the S→N acyl shift. The side chains of His736 and Asn737 are still in the vicinity of the 283-738 linkage; the Gly283 O atom close to the Asn737 O, and the His736 Nδ1 atom to the Asn737 O (see Fig. 7).

6.3. Comparison of the splicing junctions. The X10-series recombinants show notable large differences in the structure of the N- and C-extein residues, 280–283 and 739–741, to the VMA29 VDE structure in which Gly283 with the bound zinc ion near Cys 738 shows a distorted bond angle at the Ca atom and a distorted inverse-γγ turn at the scissile peptide bond to Ala284. It is reported that the VMA1 product splices correctly even in the presence of the chelating reagent, and that the addition of zinc ion rather inhibits the protein splicing in the case of the RecA intein. Therefore, the X10SSS structure with the β-strand-conformed Ser284 side-chain in contact with the Ser738 side-chain is more suitable for understanding the splicing reaction.

As for the structural analyses of the splicing reaction, the crystal structures were reported for Mycobacterium xenopi GyrA and Hedgehog autoprocessing domain. The VDE residues Cys284, His736, Asn737, and Cys738 that have been reported to be essential for the splicing reaction are superposable on the corresponding residues of the GyrA and Hedgehog proteins, having common side-chain conformations.

6.4. Scissile bond structure triggering the first reaction step. The first step involves the chemically unfavorable N→S acyl shift that breaks an amide bond and forms a high-energy thioester linkage at the N-terminal scissile bond, and hence raises a still unresolved issue of driving forces. The
GyrA crystal structure contains an extra Ala(−1) residue introduced immediately precedent to the conserved N-terminal Ser1. The peptide bond between Ala(−1) and Ser1, however, is in an energetically unfavorable cis conformation, suggesting that the cis conformation would destabilize the bond and hence generate a driving force for the first step of the reaction. The NMR structure of a splicing-active precursor of the GyrA intein showed that the amide of the Phe(−1)-Cys1 scissile bond is highly polarized and possibly in a distorted state, providing additional support for the idea that the first step in protein splicing is facilitated by destabilizing the scissile bond.

In the X10SSS structure, the trans conformation for the Gly283-Ser284 scissile bond is restrained by the contact between the Gly283 main-chain O and Ser738 side-chain O atoms, and also by the hydrogen bond between the Ser284 main-chain N and Asn359 side-chain O atoms (Fig. 5). This contact is well suited for the formation of the thiazolidine intermediate in the N→S acyl shift (see Fig. 8b). Therefore, the highly distorted conformation seems to be not essential for the reaction in the VMA1 translational protein that has the small and flexible Gly283 residue in the scissile bond, although it is not precluded that destabilizing conformations indiscernible in the X10SSS structure might transiently take place in the actual reaction.

### 6.5. Structural roles of Val403 and Ser639 in protein splicing

Splicing-defective mutants of the GST-Vma1 fusion protein, with a replacement of either Gly283Val, Cys284Tyr, His632Leu, Val403Asp, Ser639Pro, Asn737Ser, or Asn737Lys, have been identified from the random mutagenesis study. The crystal structures clearly indicate that the replacements in the key residues of Cys284, His632, and Asn737 have direct effects on the splicing reaction (see Fig. 7). The Gly283Val replacement causes a steric hindrance with the catalytic His632 side-chain. The hydrophobic side-chain of Val403 located at about 19Å from Cys284 is in interaction with that of Ile442 which is located in an anti-parallel β-sheet formed by the Tyr437-Glu443 and Ghu363-Pro369 strands. The Val403Asp replacement brings a polar side-chain into this hydrophobic environment and affects the spatial arrangement of Ile442 which is in interaction with His632. The Ser639Pro replacement abolishes a hydrogen bond between the Ser639 N and Ser635 O atoms, and disrupts the helix formation at His626-Leu640, hindering the folding of the domain II.

### 7. Chemical mechanism of the splicing reaction

Based on the crystal structures of the X10SSS precursor and X10NS spliced products, we proposed the chemical mechanism of the splicing reaction, which is summarized in Fig. 8. The X10-series VDE recombinants show that the activity of the splicing is retained by the Cys284Ser and Cys738Ser replacements, as clearly indicated by the X10NS spliced products (Fig. 6).

#### 7.1. Initiation of the N→S acyl shift

In the close proximity of the Ser284 Oγ atom of the X10SSS precursor, the Gly283 O and Ser738 Oγ atoms are located (Fig. 5). The N→S acyl shift was reported even for the recombinant with the Cys738Ala replacement. This indicates that the Gly283 O atom as an anionic form, rather than the Cys738 Sγ atom, is able to abstract a proton from the Cys284 side-chain and activate its Sγ atom in the initial step of the N→S acyl shift as exemplified in Fig. 8a.

#### 7.2. Thiazolidine intermediate

The resultant Cys284 Sγ thiolate nucleophilically attacks on the Gly283 C atom, forming an intermediate which has a five-membered ring structure, called as thiazolidine, at the Gly283-Cys284 junction (Fig. 8b). The thiazolidine ring from the reported crystal structure of its derivative is nicely superposable on the junction. The Gly283 C atom embedded in the thiazolidine ring would have a tetrahedral configuration with a hydroxy Gly283 O atom which is protonated upon the Cys284 Sγ activation. When the Cys284 Sγ atom is activated by the other catalytic atom such as the Cys738 Sγ atom, the Gly283 C atom would have an oxyanion Gly283 O atom instead of the hydroxy group, as discussed by Paulus.

The His362 side-chain is located in the vicinity of the 283-284 peptide bond. Though His362 was identified by the leucine replacement as the essential residue for the splicing reaction the X10CNS bearing the His362Asn replacement is proved to retain the splicing activity. When an imidazole ring is substituted for the amide group of the X10NS Asn362 side-chain (Figs. 5 and 7), the resultant His362 Nδ1 atom comes close to the Ser284 N atom. Therefore, the imidazole ring of His362 is so located as to stabilize the hydroxy or oxyanion group of the
thiazolidine intermediate and then as to protonate the Cys284 N atom (Fig. 8b).

7.3. Thioester intermediate. The protonation to the Cys284 N promotes the breakdown of the thiazolidine ring possibly stabilized by the His362 side-chain, and the thiazolidine intermediate is resolved into the Cys284 amino group and the Gly283-Cys284 thioester linkage toward the thioester intermediate (Fig. 8c). The reaction steps proposed here for this N→S acyl shift are entirely driven by protonation and deprotonation, in which proton ingress and egress are balanced within the splicing site.

7.4. Tetrahedral intermediate via transesterification. After the formation of the thioester intermediate, its acyl group is to be directed to the Cys738 side-chain, and its linkage with the Cys284 side-chain is to be transferred to the Cys738 side-chain. In this transesterification step, the protonated Cys284 amino-group that is formed by the opening of the thiazolidine ring activates the Cys738 Sγ.
atom into another nucleophile, which then attacks on the Gly283 thioester carbonyl carbon. Since the X10-series recombinants supposedly retain common precursory and intermediate structures, spatial arrangements of the splicing-site residues are essential also for the transsplicing. The atomic contact between the O γ atoms of the replaced Ser284 and Ser738 provides a structural basis for chemical activity of the Cys738 thiol side-chain in the course of the transsplicing that forms the second tetrahedral configuration of an oxybissulfanyl C atom at Gly283 (Fig. 8d).

7.5. Branched intermediate and succinimide formation. The resultant tetrahedral intermediate of the thioester transfer is resolved into the Cys284 thiolate and the branched intermediate\cite{13} in which the N-extein acyl group is bonded to the Cys738 side-chain with the thioester linkage (Fig. 8e). The extein regions of this branched intermediate are then cleaved off from the intein region upon the succinimide formation at Asn737 (Fig. 8f).

7.6. Final S → N acyl shift via the second thioester and thiazolidine intermediates. In the final S → N acyl shift step of the protein splicing, the thioester structure from the branched intermediate that is resolved by the hydrolysis of the succinimide (Fig. 8g) is further transformed into a peptide-bond linkage between Gly283 and Cys738. As the thioester bond thus formed is chemically unstable, the S → N acyl shift reaction would proceed more rapidly in the hydrophobic environment of the splicing site, and the spontaneous hydrolysis be prevented by the balanced deprotonation and protonation as in the initial N → S acyl shift. In the beginning of the S → N acyl shift, the neutral Cys738 amino-group formed by the succinimide hydrolysis donates its proton to the carbonyl oxygen of Asn737 in the close vicinity of the side chain of essential His736, and the thus-formed Cys738 N nucleophile attacks on the Gly283 C atom. The resultant intermediate with the tetrahedral Gly283 C (Fig. 8h) has again a thiazolidine ring structure as is formed in the initial N → S acyl shift. The Ser284 N atom of the X10SNS product is closely located to the Gly283 O of the N-extein segment so that an oxyanion of this second thiazolidine intermediate is stabilized by the Cys284 N. The thiazolidine intermediate is finally resolved into a Cys738 thiol and the peptide bond between the N- and C-exterins, releasing the splicing product of the ligated exterins (Figs. 6 and 8i).

Ludwig et al.\cite{34} have recently detected a thiazoline ring formation via a thiazolidine intermediate when the splicing assay using an artificial semi-synthetic protein trans-splicing system was carried out with a mutated intein in which the thioester formation step was blocked. This detection justifies our proposal stating the thiazolidine formation in the course of the initial N → S acyl shift (Fig. 8b). The final S → N acyl shift also involves the thiazolidine formation (Fig. 8h), and is the reverse reaction of the justified N → S acyl shift step.

8. Conclusion

Protein splicing is a novel biological reaction that implements the editing of genetic information transfer at a protein level and proceeds autocatalytically in a single molecule of precursor polypeptides. Based on the crystal structures we proposed and highlighted that the formation of the thiazolidine intermediate is a key step for the first and final acyl shifts, both of which are entirely driven by protonation and deprotonation within the splicing sites. This proposal has directed further theoretical studies with the use of ab initio quantum molecular dynamics.

The work conducted on protein splicing has already opened a new era for investigating the origin and evolutionary past of the selfish genetic element VDE in chromosomes as well as for providing novel technologies in protein chemistry.\cite{26}

We conclude that protein splicing requires catalytic residues Cys284, His362 and Asn737 as the cis-acting elements whereas Gly283 and Cys738 as the trans effectors. In addition, the short antiparallel β-sheet formed between the N-extein residues 281–283 and the C-extein residues 739–741 plays a key structural role in stabilizing the splicing junctions. Val403 is located at the long distance from Cys284 in the splicing site, and its nonconservative mutation to aspartate hinders protein splicing by causing a structural perturbation on the catalytic residues. The splicing mechanism functions only when these cis elements in the parasitic intein domain combine with the trans effectors in the host extein domains of the single spliced propeptide. Our work establishes the first milestone of thiazolidine chemistry for the world of protein splicing.

Acknowledgements

The authors express our sincere appreciation to all the collaborators of this project; Drs. Ryogo
Protein splicing: from its discovery to mechanisms

Hirata, Masato Kawasaki, Satoru Nomagi, Yoshikazu Ohya and Ryuta Mizutani. This work was supported by the Scientific Research Program of the MEXT of Japan to YS and to RM, and by the Grant-in-aid for Advanced Scientific Research on Bioscience/Biotechnology Areas from the MEXT of Japan to YA.

References


(Received Aug. 10, 2009; accepted Sep. 1, 2009)

Profile

Yasuhiro Anraku, Professor Emeritus, the University of Tokyo, was a professor (1976–1997) at the Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo and a professor (1997–2005) at the Department of Biosciences, Teikyo University of Science and Technology, Yamanashi, Japan. He was born in Manchuria (July 1936) and was educated at the University of Tokyo (B.Sc., 1960 and Ph.D., 1965). He spent his postdoctoral years at the National Institutes of Health, U.S.A (1965–1967) and the Department of Biochemistry, Stanford University School of Medicine (1967–1968). He also held a joint appointment at the National Institute for Basic Biology, Okazaki (1987–1992). The Anraku Laboratory has worked on membrane biochemistry and bioenergetics especially related to the active transport systems and respiratory terminal oxidase complexes of *Escherichia coli*, and the vacuolar H^+-ATPase of *Saccharomyces cerevisiae*. Their discoveries of the *E.coli* periplasmic binding proteins, yeast vacuolar H^+-ATPase, and protein splicing catalyzed in the VMA1 intein were the important breakthroughs in biochemistry and bioenergetics. Over 250 research papers and review articles were published in prestigious journals by his group. He was a recipient of research awards from the Japanese Biochemical Society and Matsu naga Foundation, and research grants from the Human Frontier Science Program (1990–1993; 1994–1997). Dr. Anraku has a long record of professional service as an Executive Council Member of the Japanese Biochemical Society (Vice President, 1993; President-Elect, 1995; President, 1996), Secretary-General and Chairman of the National Committee for Biochemistry, Science Council of Japan (1982–2005), and a Japanese representative to the General Assembly of the International Union of Biochemistry and Molecular Biology (IUBMB, 1982–1997). As an editor, *Journal of Biochemistry, Protoplasma, Cell Structure & Function* and IUBMB *Life* have benefited from his profound editorial insights. Internationally, he was appointed to Treasurer of the Federation of Asian and Oceanian Biochemists and Molecular Biologists (FAOBMB; 1993–1995), President (1996–1998), and Past-President (1999–2000). He also served as a Member of the IUBMB Executive Committee for 1997–2003. In 2009, he was appointed to an Honorary Membership of the FAOBMB for his contribution to establish the FAOBMB Award for young scientists.
Profile

Yoshinori Satow was born in Akita prefecture, in 1949, and started his research in 1972 on the crystal structure of the protein *Streptomyces subtilisin* inhibitor, in the Graduate School of Pharmaceutical Sciences of University of Tokyo. He took his Ph. D. degree from the School in 1977, with the determination of three-dimensional structure of this protein using protein crystallography. Then he studied the crystal structure of antibody Fab McPC603 at the Laboratory Molecular Biology of NIAMDD in the U.S. National Institutes of Health, and successfully determined the hapten-liganded Fab structure. In 1981 he was an assistant professor at Photon Factory of the National Laboratory for High Energy Physics of Japanese Ministry of Education, and was later promoted to an associate professor. He developed instrumentation and methodology for use of synchrotron radiation X-rays, especially the multi-wavelength-anomalous diffraction (MAD) method for structure determination of proteins and also X-ray spectroscopy for analyses of micro-environments of materials. Since 1988, he is a professor of the Faculty of Pharmaceutical Sciences and also the Graduate School of Pharmaceutical Sciences, University of Tokyo. He has been working on three-dimensional structures of such proteins as antibodies, human enzymes, and receptors using structural biology techniques, mostly X-ray crystallography. His current interest is in three-dimensional structures and functions of human glycoproteins and membrane-bound receptors of innate immunity and signal transduction.