THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 40, Issue of October 3, pp. 39133–39142, 2003

Printed in U.S.A.

Crystal Structure of a Mini-intein Reveals a Conserved Catalytic Module Involved in Side Chain Cyclization of Asparagine during Protein Splicing*[S]

Received for publication, June 12, 2003, and in revised form, July 23, 2003
Published, JBC Papers in Press, July 23, 2003, DOI 10.1074/jbc.M306197200

Yi Ding‡§, Ming-Qun Xu¶‖, Inca Ghosh¶, Xuehui Chen‡, Sebastien Ferrandon¶, Guillaume Lesage¶, and Zihe Rao‡§**

From the ‡Laboratory of Structural Biology and the Ministry of Education Laboratory of Protein Science, School of Life Science and Engineering, Tsinghua University, Beijing 100084, People's Republic of China, ¶New England Biolabs, Inc., Beverly, Massachusetts 01915, and §Institute of Biophysics, Chinese Academy of Science, Beijing 100101, People's Republic of China

We have determined the crystal structure of a 154-residue intein derived from the *dna*B gene of *Synechocystis* sp. strain PCC6803 and refined it to a 2.0-Å resolution. The x-ray structure suggests that this intein possesses two catalytic sites that appear to be separately responsible for splicing and cleavage of the N- and C-terminal scissile bonds. The conserved intein block F residues are the important components of a catalytic site for side chain cyclization of the last intein residue, Asn-154. The data suggest that the imidazole ring of His-143 is involved in the activation of the side chain Nδ atom of Asn-154, leading to a nucleophilic attack on the carbonyl carbon of Asn-154. Substitution of His-143 with Ala or Gln resulted in the inhibition of C-terminal cleavage. His-153, Asp-136, and a water molecule appear to constitute an oxyanion binding site by contacting the carbonyl oxygen of Asn-154 to stabilize the transition state. The structure and mutagenesis data also support that the close contact between the hydroxyl groups of Thr-138 and Ser-155, whose side chain participates in an S → O acyl shift, plays an important role in the nucleophile orientation. Our structural modeling suggests that this catalytic module is conserved in the C-terminal subdomains of inteins from diverse organisms.

Protein splicing is a posttranslational processing event that involves the precise removal of an intervening sequence, an intein, from a protein precursor with concomitant ligation of the flanking protein sequences (N and C exteins) via a native peptide bond (1–3). *In vitro* splicing of inteins in heterologous proteins suggests that inteins with the first C-extein residue contain sufficient information needed for the autocatalytic splicing process without the requirement of exogenous energy or protein co-factor (4, 5).

Among the more than 130 inteins identified so far, the majority contain two discrete functional domains, the homing endonuclease domain and the splicing domain (Fig. 1*A*) (InBase, the New England Biolabs Intein Data Base (6)). The protein splicing function of the intein does not depend on their homing endonuclease activity since splicing-proficient minimal inteins (mini-inteins) occur naturally (7–9) and have been generated by deleting the centrally located endonuclease domain. The splicing domain of a dual function intein appears to be split by the endonuclease domain into two segments, the N- and C-terminal subdomains. The N-terminal subdomain (in the range of 100–150 residues) contains the conserved intein blocks A, B, N2, and N4, whereas the C-terminal subdomain of ~35–50 residues contains blocks F and G (6, 10). The N-terminal subdomain of inteins not only exhibits sequence homology with the hedgehog proteins of eucaryotes but also shares an acyl rearrangement mechanism of breaking the peptide bond preceding a nucleophile-containing residue (11). This raises the question of whether inteins share a common ancestor with other autoprocessing proteins and have evolved by acquiring a new catalytic capacity imbedded in their C-terminal subdomain.

Biochemical and mutational studies reveal that protein self-splicing consists of a series of concerted nucleophilic replacements (4, 12, 13) (Fig. 1*B*). The splicing reaction is dependent on a set of highly conserved residues at both splice junctions (12, 14–16). The vast majority of inteins possess an asparagine residue at their C termini and a hydroxyl- or thiol-containing residue after each splice junction. Splicing in these inteins is initiated by an acyl rearrangement involving the side chain of serine or cysteine at the N-terminal splice junction, resulting in a (thio)ester bond (12, 17). Next, a branched intermediate is formed as a result of transesterification by the sulfhydryl- or hydroxyl-containing side chain of cysteine, serine, or threonine at the C-terminal splice junction (5, 13). The last catalytic step involves the excision of the intein by side chain cyclization of the C-terminal asparagines (13, 18). This yields a transient (thio)ester intermediate that can undergo a spontaneous acyl rearrangement to form a native amide bond between the two exteins. It has been observed that the splicing reaction could be suppressed by mutations, often leading to cleavage at N-terminal and/or C-terminal splice junctions (12, 15, 19, 20). Therefore, it is of interest to understand how the spatial arrangement of an intein active site could account for independent scission of the peptide bonds flanking an intein.

* This research was supported by Projects 30128002 and G1999075600 (China). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Fig. 1.

The atomic coordinates and structure factors (code 1MI8) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‖ To whom correspondence may be addressed: New England Biolabs, Inc., 32 Tozer Rd., Beverly, MA 01915. Tel.: 978-927-5054; Fax: 978-921-1350; E-mail: xum@neb.com.

** To whom correspondence may be addressed. Tel.: 86-10-62771493; Fax: 86-10-62773145; E-mail: raozh@xtal.tsinghua.edu.cn.

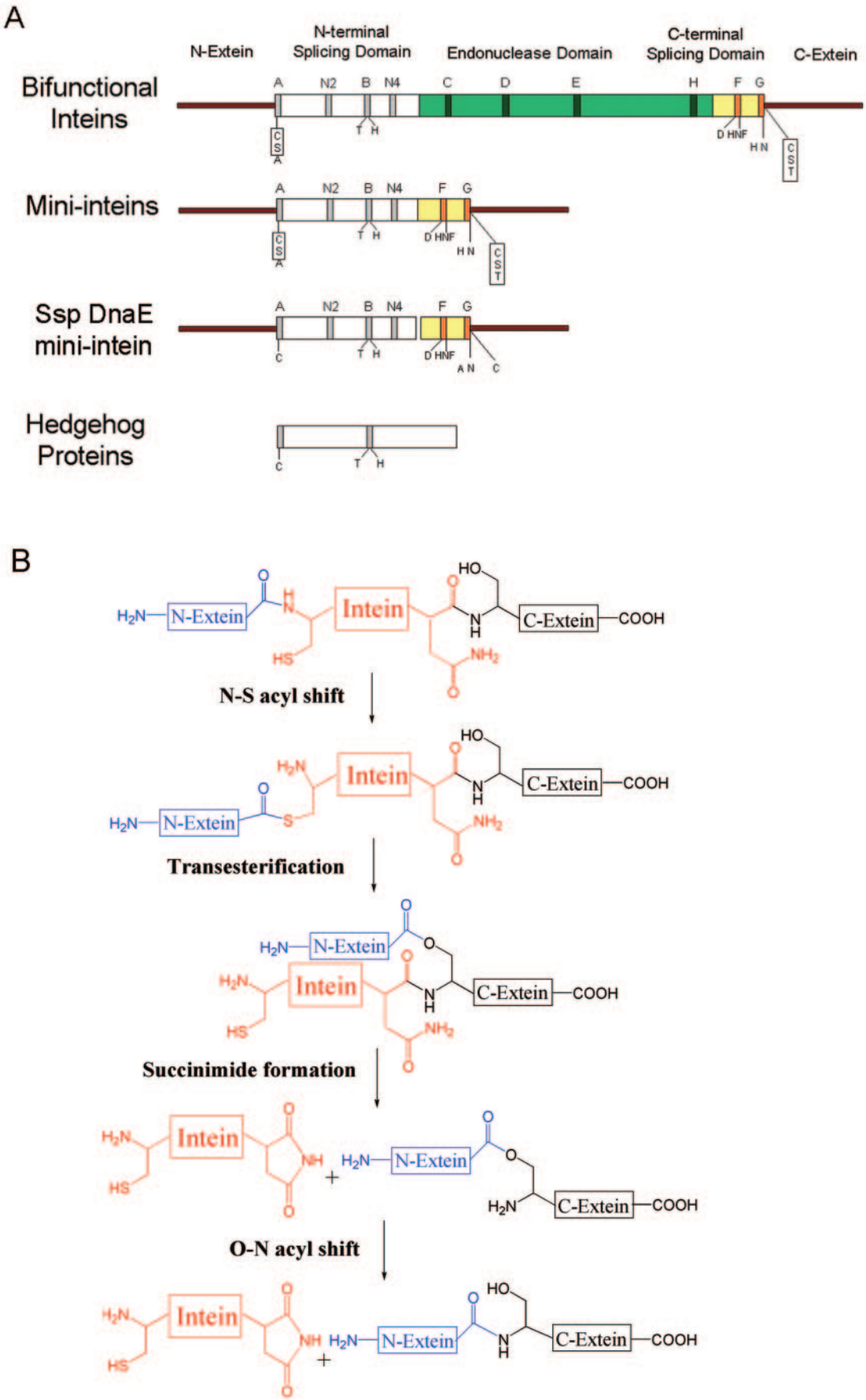

FIG. 1. **Conserved intein motifs and a 4-step splicing pathway.** *A*, diagram of bifunctional inteins, mini-inteins, the split Ssp DnaE intein, and hedgehog proteins. Blocks A, N2, B, and N4 (*gray*) in the N-terminal subdomain (*white*) and blocks F and G (*orange*) in C-terminal subdomain (*yellow*) are shared by the splicing domains; blocks C, D, E, and H are found within the endonuclease domain (*green*). Residues involved in

The x-ray structures of the intein of the vacuolar ATPase subunit from *Saccharomyces cerevisiae* (Sce VMA intein) (21–24), the archaeal PI-PfuI intein from *Pyrococcus furiosus* (PI-PfuI intein) (25), and the mini-intein of the GyrA protein from *Mycobacterium xenopi* (Mxe GyrA intein) (26) have revealed that inteins contain a horseshoe-like β-strand scaffold termed the Hint (Hedgehog, intein) module (11). Mutation of catalytic residues at the N- and C-terminal splice junctions of the Sce VMA inteins resulted in the trapping of the inactive precursor proteins for crystallization studies (22, 23). These investigations indicated that the conserved block B residues and the residues at the N-terminal splice site form an active center that facilitates cleavage of the N-terminal scissile peptide bond. In addition, it has been suggested that the penultimate histidine residue plays a key role in succinimide formation because of its proximity to the side chain of the C-terminal Asn (26). Other interactions may exist to assist the cyclization of Asn in the more than 20 inteins known to lack the penultimate histidine (27–30). Mutation of the highly conserved block F residues impaired splicing (31); however, the structural and catalytic roles of these residues have not been elucidated. Furthermore, the mechanisms for generating a nucleophilic potential at the intein N terminus and at the first C-extein position have yet to be determined. A complete examination of a pre-splicing structure would lead to further understanding of the chemical mechanism of protein splicing and the evolution pathway of autocatalytic inteins.

To explore the outstanding questions concerning the structure and evolution of inteins, we conducted a crystal structure study and mutational analysis of a mini-intein precursor. This mini-intein was derived from a 429-amino acid intein encoded in the DNA helicase gene of the cyanobacterium *Synechocystis* sp. strain PCC6803 (Ssp DnaB intein) (27, 32). Deletion of a centrally located 275-amino acid sequence encoding an endonuclease domain resulted in a splicing-proficient minimal intein (Ssp DnaB mini-intein) (27). Our investigation determined that two catalytic sites within the structure independently catalyze cleavage of the peptide bonds flanking the intein. The C-terminal catalytic site, formed by the conserved residues of blocks F and G, constitutes a charge relay system and an oxyanion binding site, which may play vital roles in catalyzing the cyclization of Asn-154. This intein structure also reveals that the peptide bond between Asn-154 and the first C-extein residue, Ser-155, is in a distorted *trans* conformation. The implications of the existence of a catalytic site in each intein subdomain are discussed with respect to intein evolution.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The pTWIN vector (33) was used for expression and purification of the Ssp DnaB mini-intein in *Escherichia coli*. Purification of the Ssp DnaB intein was conducted by an intein-mediated cleavage method, as described previously (34). The protein sample was then applied to a Resource Q column (Amersham Biosciences) and eluted with a linear gradient of NaCl. The peak containing the Ssp DnaB mini-intein was loaded onto a Superdex 75 column (Amersham Biosciences) pre-equilibrated with 10 mM Tris-HCl, 25 mM NaCl, pH 7.5, 10 mM β-mercaptoethanol and eluted with the same buffer. The final protein concentration was 24 mg/ml before crystallization. Dithiothreitol was added to the protein to a final concentration of 5 mM (35).

Mutagenesis and C-terminal Cleavage Activity Assay—The gene encoding the Ssp DnaB intein with the C1A mutation was cloned into the *Xho*I and *Age*I sites of the LITMUS 28 vector (New England Biolabs, Inc.). Kunkel mutagenesis was used to generate the point mutations in the intein using the following sense strand primers (36, 37): V134A, 5′-ACTGGAGTCGAAGAGGCTTTTGATTTGACTGTGC-3′; F135A, 5′-ACTGGAGTCGAAGAGGTTGCTGATTTGACTGTGC-3′; F135Y, 5′-ACTGGAGTCGAAGAGGTTTATGATTTGACTGTGCCA-3′; D136A, 5′-GTCGAAGAGGTTTTTGCTTTGACTGTGCCAGGA-3′; H143A, 5′-ACTGTGCCAGGACCAGCTAACTTTGTCGCGAA TG-3′; H143Q, 5′-ACTGTGCCAGGACCAGAGAACTTTGTCGCGAATG-3′; N144A, 5′-ACTGTGCCAGGACCACATGCCTTTGTCGCGAATGAC-3′; F145A, 5′-CAGGACCACATAACGCTGTCGCGAATGACATCAT-3′; F145Y, 5′-CCAGGACCACATAACTATGTCGCGAATGACATCAT-3′; N144A, 5′-ACTGTGCCAGGACCACATGCCTTTGTCGCGAATGAC-3′. The mutants were verified by DNA sequencing, and the mutated intein was transferred to the vector pMSL to replace the wild type intein sequence using *Xho*I and *Age*I (19). The pMSL vector expresses a tripartite fusion protein with maltose-binding protein (MBP)[1] at the N terminus followed by the Ssp DnaB intein (17 kDa) containing the C1A mutation and T4 DNA ligase (58 kDa). The plasmids were transformed into *E. coli* ER2566, and a colony was inoculated in LB medium containing 100 μg/ml ampicillin. The cells were grown at 37 °C until A_{600} of 0.5 was reached and induced for 3 h at 30 °C with the addition of 0.3 mM isopropyl β-D-thiogalactoside. *In vivo* cleavage activities of the various mutants were examined by Coomassie Blue staining of 12% SDS-PAGE. The identity of the precursor and cleavage products was confirmed by Western blot analysis using antibodies against MBP and the Ssp DnaB intein. Amino acid sequencing confirmed that the MI species contained the sequence expected for the N terminus of MBP (MKTEEGKLV), and the L species contained the N terminus of the expected C-extein (SIEQDTGMLP). The T138A or T138S mutants were examined in a construct (pMIB) in which the mutated DnaB intein sequence was inserted between the *E. coli* MBP and the chitinase A1 chitin binding domain (CBD) of *Bacillus circulans* (31). Expression of the wild type or mutant MIB fusion proteins was induced at 37 °C for 3 h. *In vivo* splicing activity was analyzed by Coomassie Blue staining of a 10–20% SDS-PAGE. The identity of the splicing and cleavage products was confirmed by Western blot analysis with antibodies against MBP, the DnaB intein, and CBD.

Crystallization and Data Collection—Crystals were grown in hanging drops equilibrated by vapor diffusion against reservoir solutions of 100 mM Tris buffer at pH 7.5–8.0, 16–20% polyethylene glycol 4000, and 4% (v/v) ethyl glycerol (35). The crystals were mounted on a nylon-fiber loop and flash-frozen in a nitrogen-gas stream at 100 K before x-ray data analysis. Diffraction data were collected using a Mar345 Image Plate with a Rigaku rotating copper anode x-ray generator at 48 kV and 98 mA (λ = 1.5418 Å). All intensity data were processed and scaled using the programs DENZO and SCALEPACK (38).

Structure Determination and Refinement—The structure was determined by molecular replacement using AMORE from the CCP4 software package. The crystal structure of the GyrA intein from *M. xenopi* (PDB entry 1AM2) was used as the search model. The amino acid sequence similarity between the Mxe GyrA intein and the Ssp DnaB mini-intein is 41%. The two α-helices were omitted from the model, and the nonhomologous residues in the model were substituted with alanine. The rotation solution was obtained using 8.0–4.0-Å data, and the highest peak appeared with $\alpha = 2.32°$, $\beta = 105.75°$, $\gamma = 141.58°$. Translation searches were carried out in the two possible space groups $P3_121$ and $P3_221$, the latter giving a better solution ($x = 0.3328$, $y = 0.3246$, $z = 0.2620$), which was significantly above the noise level. The correlation coefficient and R-value were 0.186 and 0.519, respectively. After rigid body refinement, the R-factor was reduced to 0.496.

The model was rebuilt with O (39) and refined with CNS (40). No data truncation was applied in the refinement. 10% of the data was set aside to calculate the free R-factor. Because of the low sequence similarity between the GyrA intein and Ssp DnaB intein, significant model building work was done manually using O involving mutations, deletions, and insertions. Using the data of 40–2.0 Å, $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps were calculated and examined with O. The refinement was completed by alternating between manual building and minimization using data in the resolution range 40–2.0 Å. Group

[1] The abbreviations used are: MBP, maltose-binding protein; I, intein; MI, MBP intein; CBD, chitin binding domain.

nucleophilic attack (*letters* in *box*) as well as other highly conserved amino acids are indicated *below* the block diagram. The N-terminal subdomain shares sequence similarity with hedgehog proteins of eucaryotes. Conserved residues between inteins and hedgehog proteins are shown *below* the hedgehog protein block diagram. *B*, schematic representation of the four-step splicing pathway by the Ssp DnaB intein. The intein (*red*) is excised from the precursor protein coupled with the ligation of N-extein (*blue*) and C-extein (*black*). The mechanism is as originally proposed (12).

TABLE I
Diffraction data collection and structure refinement statistics

Space group	$P3_221$
Unit-cell dimensions (Å)	$a = b = 58.2$ $c = 70.3$
Resolution range (Å)	40.0-2.0 Å
Number of observations	107,903
Number of unique reflections	9,307
Completeness (%)	97.6 (75.5)[a]
Average $I/\sigma(I)$	20.5 (7.5)[a]
R_{merge}	0.046 (0.262)[a]
Mean redundancy	11.6
Number of protein molecules per asymmetric unit	1
Number of protein atoms	1,107
Number of water molecules	52
R_{work}	0.210
R_{free}[b]	0.263
Root mean square deviations from idealities	
Bond lengths (Å)	0.013
Bond angles (degree)	2.0
Average B factors (Å^2)	
Protein	30.2
Waters	33.4

[a] Numbers in parentheses are for the resolution range 2.07-2.00 Å.
[b] R_{free} is calculated for the 10% subset of the unique reflections.

B-factor refinement was then used to refine the temperature factors in the model. In each refinement step, initial anisotropic overall B-factor correction with a lower resolution limit of 6 Å and a bulk-solvent correction were applied to the data. The 52 water molecules were located in difference electron density peaks only if the peaks were above 3.0 σ and if acceptable hydrogen bonds to atoms in the model could be made. The final protein model and the waters were checked using O. The highly ordered water molecule, Wat-30, carries a B-factor of ~26 Å^2, lower than the average of solvent molecules. The model was refined to an R of 0.210 and an R_{free} of 0.263 at 2.0-Å resolution. No residues are in the disallowed regions of the main-chain φ-ψ plot. Three N-terminal extein residues (from −5 to −3), three C-terminal extein residues (from 157 to 159), and the loop between residues 99 and 115 were found to be disordered and have been omitted from the model.

A model of the wild type DnaB mini-intein precursor was generated by replacing Ala-1 and Ala-154 in the determined intein structure with Cys and Asn, respectively. The model was imported into the CNS (40) program for energy minimization. The Cys-1 and Asn-154 side chain atoms were initially modeled with their χ_1 angles equal to the most favored conformation, Cys $\chi_1 = -60°$ and Asn $\chi_1 = -180°$. After energy minimization, the angles were only slightly changed. The backbone angle N-Cα-C (τ) at the C-terminal scissile bond (Ser-155) measures 122° both before and after energy minimization.

RESULTS AND DISCUSSION

Identification of Two Catalytic Sites—The Ssp DnaB mini-intein is comprised of an N-terminal segment of 106 residues and a C-terminal segment of 48 residues (27). The precursor protein used in this study possesses five native N and C extein residues flanking the 154-residue mini-intein. The first and last intein residues (Cys-1 and Asn-154) were substituted with alanine to capture the precursor in a presplicing state. The crystal structure of this intein precursor was determined by the molecular replacement method and refined to 2.0 Å resolution. The statistics of diffraction data collection and structure refinement are given in Table I. The structure exhibits a typical horseshoe-like 12 β-strand Hint domain (Fig. 2*A*). The N-terminal extein residues Ser-2 and Gly-1, and C-terminal extein residues Ser-155 and Ile-156 have well defined electron density maps (Fig. 2, *B* and *C*). The N- and C-terminal intein subdomains are tightly associated, with numerous hydrogen-bonding contacts and hydrophobic packing interactions. Superposition of the available crystal structures indicates that inteins share overall structural similarity. The root mean square difference for the Cα atoms of the Hint domain is ~1.4 Å between the Ssp DnaB and Sce VMA inteins (21–24), 1.2 Å between the Ssp DnaB and Mxe GyrA inteins (26), and 1.2 Å between the Ssp DnaB and *P. furiosus* Pol inteins (25).

A model of the wild type DnaB mini-intein precursor was generated by replacing Ala-1 and Ala-154 in the determined intein structure with Cys and Asn, respectively. The two splice junctions are located at the center of the horseshoe structure at the ends of two adjacent antiparallel β-strands. Intriguingly, the modeled structure precursor reveals the presence of two catalytic sites that are in close proximity to each other but which do not share catalytic components (Figs. 2 and 3). Several residues in the Ssp DnaB intein N-terminal region appear to play major roles in aligning and activating the reactive groups at the N-terminal splice junction (Fig. 2*D*). The important residues include Gly-1 (the native extein residue preceding the scissile bond) and Cys-1, flanking the N-terminal splice site, Thr-70, Asn-72, and His-73 of block B, and Thr-51. Contributions from the highly conserved block B residues have been observed in the x-ray structures of the Sce VMA and Mxe GyrA inteins (22, 23, 26). The S-H bond of the modeled Cys-1 is polarized by hydrogen bonding with the hydroxyl group of Thr-51 and the carbonyl oxygen of Gly-1. The conformation of Thr-51 appears to be maintained in the appropriate orientation for hydrogen bonding by interactions with Lys-54 and Val-134. The arrangement of this active site is capable of facilitating the initial step in the splicing pathway, a nucleophilic attack by the thiol of Cys-1 on the carbonyl carbon of the peptide bond between Cys-1 and Gly-1 to form a thioester intermediate (Fig. 4*A*).

The second catalytic site is formed by the conserved residues in block F and G in the intein C-terminal subdomain and surrounds the Asn-154 and Ser-155 that are involved in two crucial steps of the splicing reaction (Figs. 2*E*, 3, and 4). This site may be responsible for enhancing the nucleophilicity of the Asn-154 side chain and stabilizing the resulting tetrahedral intermediate. In addition, the scissile bonds at the N- and C-terminal splice junctions are in the ordinary *trans* conformation. The backbone angle N-Cα-C (τ) at the C-terminal scissile bond (Ser-155) measures 122°, deviating from the ideal angle of 110° by 12°. The implications of these findings are discussed in the next sections.

Structure and Function of the C-terminal Catalytic Site—This study provides the first structural evidence that several highly conserved block F and G residues, located in neighboring antiparallel β strands (β10, β11, and β12), play important roles in the formation of a splicing intermediate via cyclization of the Asn-154 side chain (Figs. 2–4). First, the imidazole side chain of His-143 is within hydrogen-bonding distance with a highly ordered water molecule (Wat-30), which in turn is in hydrogen-bonding contact with the Nδ atom of the Asn-154 side chain. We propose that Asn-154, His-143, and Wat-30 constitute a charge relay system whereby the imidazole side chain of His-143 deprotonates Wat-30, which in turn accepts a proton from the Nδ atom of Asn-154. His-143 and Asn-154 also form hydrogen bonds with their neighboring residues, which may contribute to the conformational stability of His-143 and Asn-154 (Fig. 3). The highly ordered water molecule (Wat-30) is also stabilized by hydrogen bonding with Val-139 (Fig. 3). The role of His-143 in C-terminal cleavage was further investigated by mutagenesis, as described in the next section.

The τ angle (N-Cα-C) of Ser-155 measures 122°, deviating ~12° from the ideal 110° along the main chain dihedral angle distortion, indicating that the scissile bond at the C-terminal splice junction is in a high energy conformation. The presence of such a labile peptide bond lowers the energy barrier to nucleophilic attack by the side chain of Asn-154. The interactions of Ser-155 with Leu-137 and Ile-156 within the active site may contribute to the retention of the labile conformation of this residue. The N- and C-terminal scissile bonds of a Sce

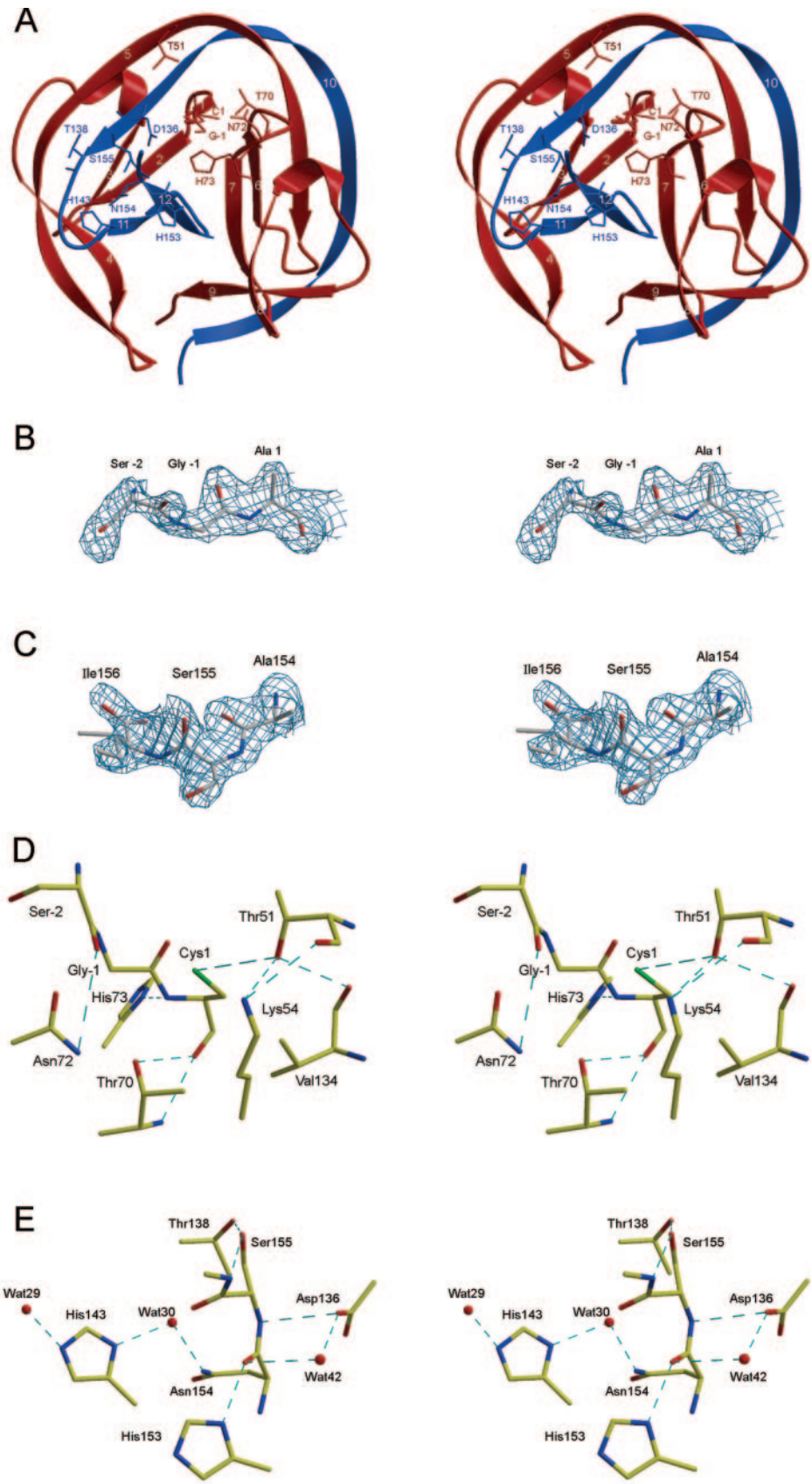

FIG. 2. **Stereo view of the Ssp DnaB mini-intein precursor structure.** *A*, stereo ribbon drawing of the Cα backbone and the catalytic residues of the Ssp DnaB intein. The β-strands are *numbered* on the *ribbon*. The N-terminal intein sequence is shown in *red*, whereas the C-terminal intein sequence is shown in *blue*. The important residues involved in catalysis are shown in a *ball-and-stick* representation. *B*, stereo view of the N-terminal splicing junction showing the first intein residue and the extein residues adjacent to it. (omit map, contoured at 1σ for the 2.0 Å resolution structure). *C*, stereo view of the C-terminal splicing junction showing the last intein residue and the extein residues adjacent to it (omit map, contoured at 1σ for the 2.0 Å resolution structure). *D*, close-up stereo view of the catalytic module of the N-terminal subdomain. The residues are shown in a *ball-and-stick* representation. Carbon atoms are colored *yellow*, oxygen atoms are *red*, nitrogen atoms are *blue*, and sulfur atoms are *green*. Hydrogen bonds are shown by *cyan dashed lines*. *E*, close-up stereo view of the catalytic site of the C-terminal subdomain. For clarity, some obscuring residues and related interactions are omitted. The residues are shown in a *ball-and-stick* representation. Colors are as described in *B*. These figures were drawn using BOBSCRIPT (49) and RASTER3D (50).

VMA intein precursor were previously found in distorted *trans* conformations (22, 23), whereas the scissile bond at the N-terminal splice junction of the Mxe GyrA intein, possessing one N-extein residue and no C-extein residue, was reported in a *cis* conformation (26). This dissimilarity could be because of the different interactions within each protein sequence, resulting in the trapping of the scissile bonds in different conformations.

The carboxylate group of Asp-136 is hydrogen-bonded to a water molecule (Wat-42), which in turn is in hydrogen-bonding distance with the carbonyl oxygen of Asn-154. The imidazole

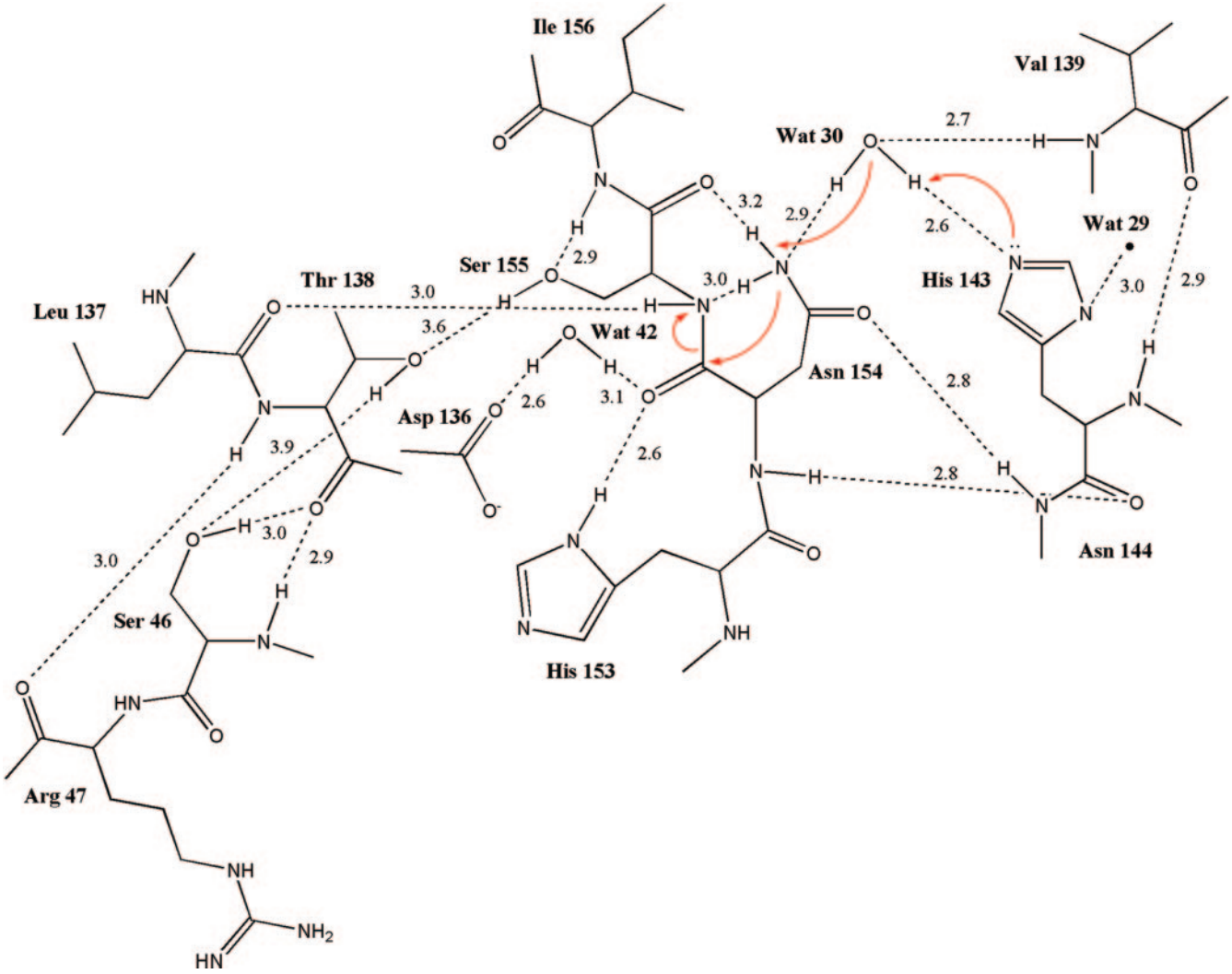

FIG. 3. **Diagram of the residues present in the C-terminal catalytic site of the Ssp DnaB intein.** *Dashed lines* indicate hydrogen bonds, and *numbers* are distances in angstroms. The *red arrows* indicate the routes of nucleophilic attacks in the splicing pathway. The energy-minimized wild type intein model was used to generate this illustration.

ring of His-153 is also hydrogen-bonded with the carbonyl oxygen of Asn-154. We propose that, upon tetrahedral intermediate formation, the now negatively charged carbonyl oxygen of Asn-154 is stabilized by hydrogen bonding with the imidazole ring of His-153 and Wat-42, which in turn contacts the carboxylate group of Asp-136 (Fig. 4*D*). Collapse of the tetrahedral intermediate is facilitated by the transfer of a proton to the leaving amide group from Wat30, which in turn accepts a proton from the imidazolium side chain of His-143. A spontaneous O → N acyl shift results in a stable peptide bond between Gly-1 and Ser-155 (Fig. 4*E*). This structural study indicates that two highly conserved histidine residues, His-143 and His-153, of the C-terminal subdomain play distinct roles in catalyzing Asn cyclization, contrary to previous suggestion that the penultimate histidine directly participates in the activation of Asn side chain (21, 26).

The nucleophilic attack by the first C-extein residue (Ser-155) on the carbonyl carbon of the thioester bond at the N terminus of the intein (Fig. 4*B*) results in the formation of a branched ester intermediate. It is, therefore, of interest to examine the interactions that affect the seryl Oγ for an S → O acyl transfer reaction. A residue such as a histidine, which can participate in general base catalysis, is not observed contacting Ser-155. However, in non-catalyzed reactions rapid intramolecular S → O acyl transfer has been observed when a hydroxyl group is present in close proximity to a thioester (41). The structure suggests that the orientation of the Ser-155 Oγ may be strongly influenced by the close contact with the hydroxyl group of Thr-138 (Fig. 4*B*). The conformation of Thr-138 is also stabilized by interactions with Ser-46 and Arg-47. These contacts can contribute to splicing by positioning the hydroxyl group of Ser-155 close to the N-terminal thioester.

Interestingly, the distance between the Sγ atom of the modeled side chain of Cys-1 and the hydroxyl group of Ser-155 in the DnaB intein precursor is ~8.5 Å (Fig. 4*A*). This observation is similar to the finding in the Sce VMA intein precursor (containing the C1A and N454A mutations), which shows a ~9-Å distance between the N- and C-terminal reacting groups (22). These distances are longer than the 3.6 Å between the N and C termini of the Sce VMA intein precursor, which was trapped by a different set of amino acid substitutions (Cys-1, Asn-454, and Cys-455 were all substituted with serine) (23). A conformational change may have to occur to traverse this relatively large gap after the initial acyl shift step.

Mutagenesis Study of Block F Residues—Site-directed mutagenesis was performed to determine the role of highly conserved residues in intein block F on C-terminal cleavage activity. Single amino acid substitutions were introduced in a Ssp DnaB intein sequence carrying a C1A mutation, which blocks cleavage at the upstream splice junction (Fig. 5). Precursors

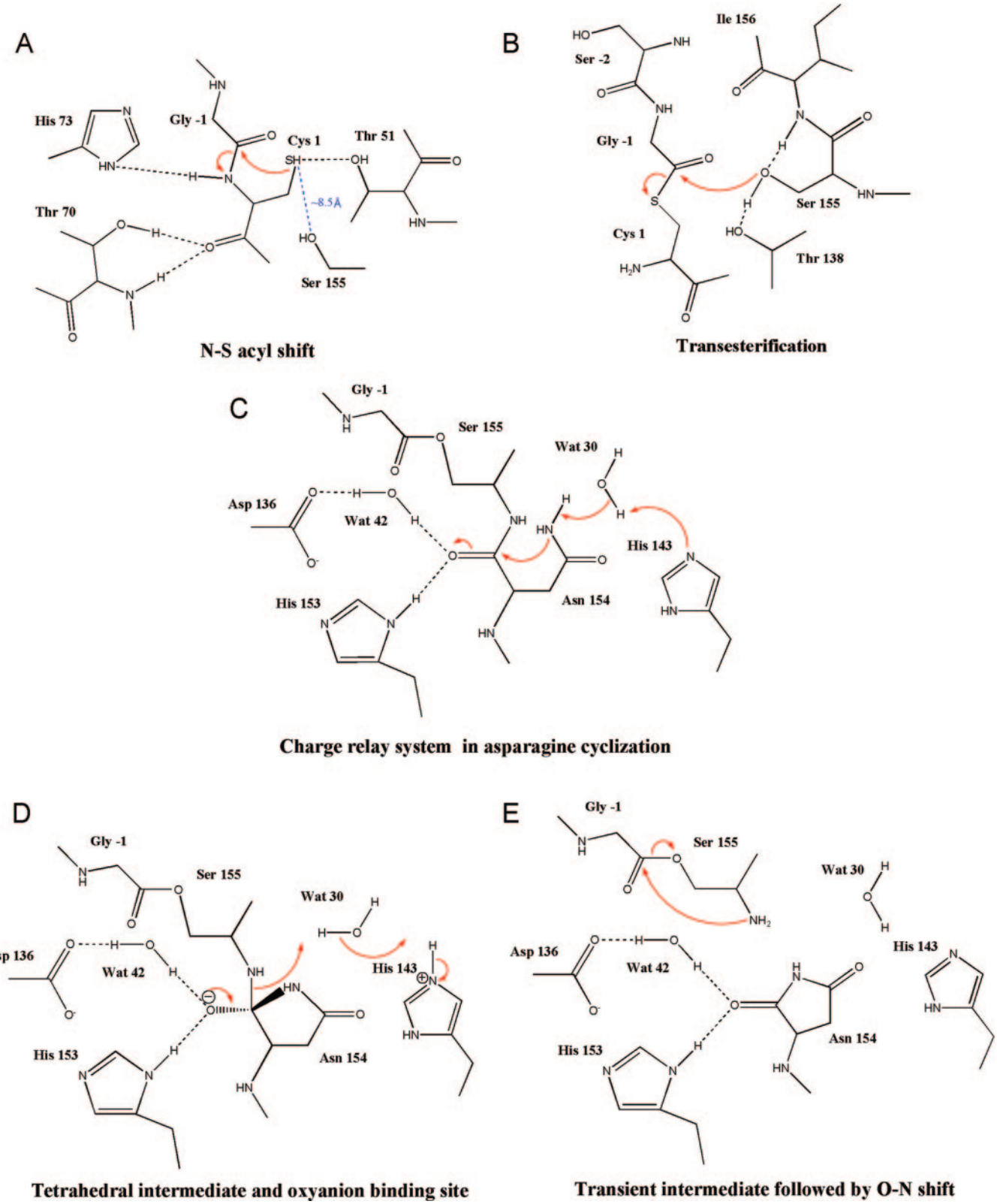

FIG. 4. **A chemical mechanism proposed for splicing of the Ssp DnaB intein.** The *red arrows* indicate the routes of nucleophilic attacks in the splicing pathway. *Dashed lines* indicate hydrogen bonds. The tetrahedral intermediate formed by an N-S acyl rearrangement at Cys-1 is not shown.

were expressed as tripartite fusion proteins, with the intein (I) fused at its N and C termini to MBP as the N-extein (M, 42 kDa) and T4 DNA ligase as the C-extein (L, 58 kDa). Examination of the *E. coli* cell extracts on SDS-PAGE indicated that complete scission of the peptide bond at the C-terminal splice junction occurred in the control containing the C1A substitution, generating a 64-kDa MBP-intein (MI) and a 58-kDa C-extein domain. The H143A substitution partially impaired C-terminal cleavage, whereas the H143Q mutation completely blocked cleavage, resulting in the accumulation of the precursor species. The identity of the precursor and cleavage products was verified by Western blot analysis with antibodies against

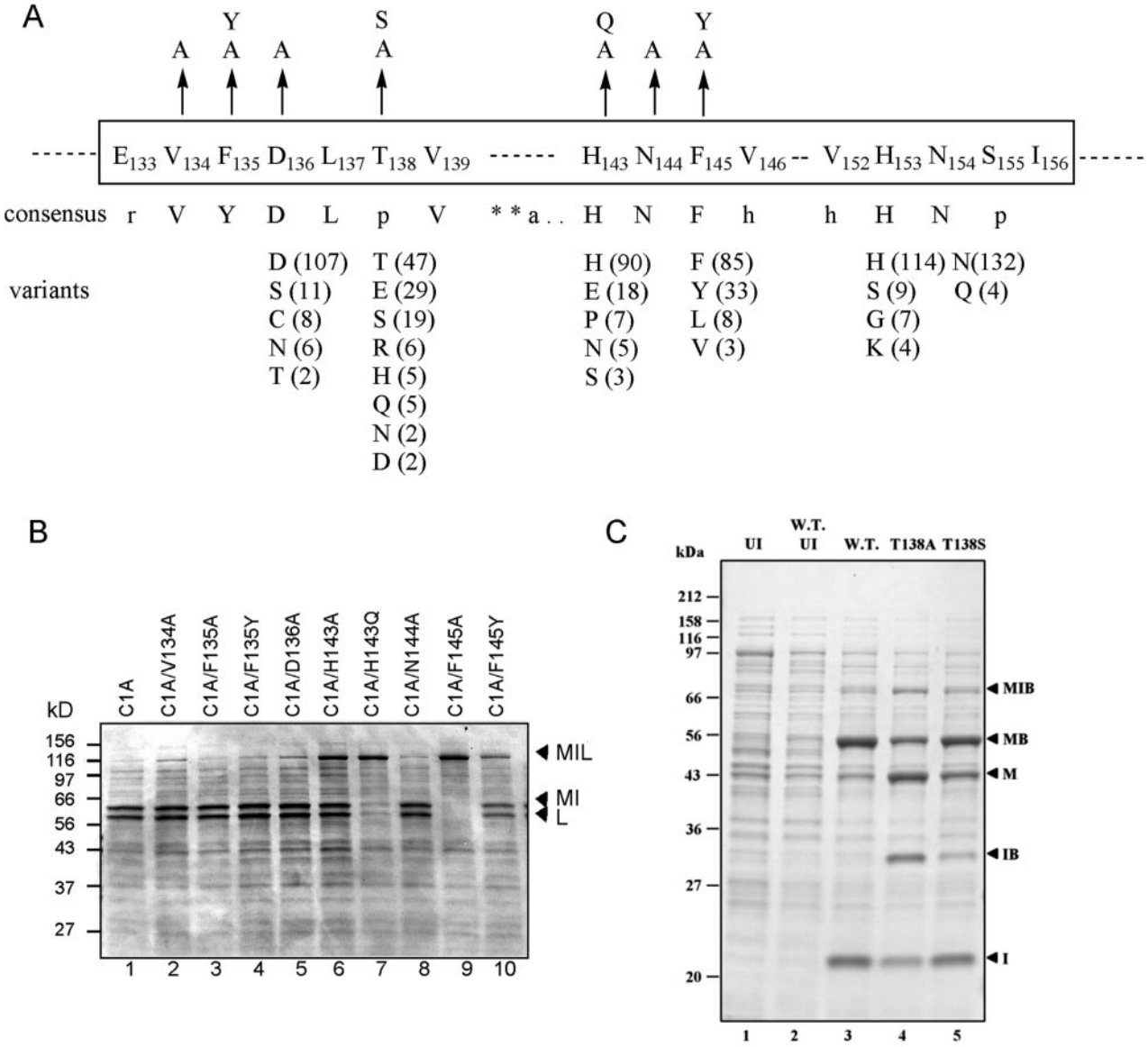

FIG. 5. **Mutagenesis study on the conserved block F residues and their role in Asn cyclization.** *A*, conserved residues in the C-terminal subdomain of the Ssp DnaB intein are shown in a *box*. The amino acid substitutions are indicated by *arrows*. The consensus sequence (the *first line below* the *box*) is from InBase, the New England Biolabs Intein Data Base (6). In the consensus line key, *r* indicates aromatic residues (*F*, *Y*, *W*); *p* indicates polar residues (*S*, *T*, *E*); *a* indicates acidic residues (*D*, *E*), and *h* indicates hydrophobic residues (*V*, *L*, *I*, *M*, *A*, *G*). The variants of the 136 inteins present in the data base at the time of our comparison are shown *below* the *box*, and the numbers of inteins possessing the amino acid at the specific position are shown in *parentheses*. For simplicity, the residues present in low frequency are omitted. *B*, effect of block F residue mutagenesis on C-terminal cleavage activity. The mutants were expressed as tripartite fusion proteins (*MIL*) with the intein (*I*) fused at its N and C termini to MBP as the N-extein (*M*, 42 kDa) and T4 DNA ligase as the C-extein (*L*, 58 kDa). Scission of the peptide bond at the C-terminal splice junction, as shown in the control carrying the C1A substitution (*lane 1*), generated a 64-kDa MBP-intein (MI) and a 58-kDa C-extein domain (L). The cleavage activities of the mutants (as indicated above the gel) were examined by Coomassie Blue-stained SDS-PAGE; *lane 2*, C1A/V134A; *lane 3*, C1A/F135A; *lane 4*, C1A/F135Y; *lane 5*, C1A/D136A; *lane 6*, C1A/H143A; *lane 7*, C1A/H143Q; *lane 8*, C1A/N144A; *lane 9*, C1A/F145A; *lane 10*, C1A/F145Y. *C*, the effect of the T138A or T138S mutation on splicing and cleavage activity. The wild type or mutated intein sequence (I) was inserted between the *E. coli* maltose-binding protein (*M*, 43 kDa) and the chitinase A1 chitin binding domain (*B*, 7 kDa) of *B. circulans* (31). Expression of the MIB fusion proteins was induced at 37 °C for 3 h. The cell extracts were analyzed by Coomassie Blue staining of a 10–20% SDS-PAGE; *lane 1*, uninduced cell extract carrying no plasmid; *lane 2*, uninduced wild type (*W.T.*) construct; *lane 3*, wild type construct; *lane 4*, T138A mutant; *lane 5*, T138S mutant. The Ssp DnaB intein (I) exhibited an apparent molecular mass of 22 kDa on SDS-PAGE. Splicing resulted in the spliced product (MB, 50 kDa) and the intein (I), whereas cleavage at the N-terminal splice junction yielded M and IB (29 kDa). The identity of the splicing and cleavage products was confirmed by Western blot analysis using antibodies against MBP, the Ssp DnaB intein, and CBD (data not shown).

MBP and the Ssp DnaB intein and further confirmed by amino acid sequencing (data not shown). Replacement of Phe-145 with alanine dramatically reduced C-terminal cleavage, whereas a conservative replacement with Tyr only partially impaired cleavage activity (Fig. 5). Phe-145 makes hydrophobic contacts with the conserved residues Leu-137 and Val-139 in block F and Val-152 in block G as well as with Ile-2, Ile-8, and Ile-30 in the N-terminal subdomain. Any perturbations to the hydrophobic interactions will affect the conformation of the active site and the precise alignment of the splicing junctions, thus impacting splicing or cleavage. Because Phe-145 and His-143 are located on the β_{11} strand, a mutation of Phe-145 could affect the position of His-143. Replacement of Val-134, Phe-135, Asp-136, or Asn-144 with Ala showed only a slight effect on C-terminal cleavage activity, resulting in marginally more precursor species compared with the control sample, a C1A

mutant (Fig. 5). The data indicate that His-143 plays a significant role in catalyzing cyclization of the side chain of Asn-154. The N136A mutation had only a marginal effect on the cleavage activity of the Ssp DnaB intein. This is probably because of the low sensitivity of the *in vivo* assay system but can also be explained by the complementary effect of His-153 present in the Ssp DnaB intein.

To investigate whether Thr-138 plays a role in facilitating splicing, possibly by interacting with the side chain of Ser-155, the effect of mutation of Thr-138 on splicing was examined by using a construct (MIB) in which the DnaB intein carrying a T138A or T138S mutation was inserted between MBP and CBD. The T138A mutant exhibited significantly reduced splicing when the expression of the MIB fusion protein was induced at 37 °C, yielding considerably less spliced product, MBP-CBD (MB) and the excised intein (I) (Fig. 5*C*). The T138A mutant precursor underwent predominantly N-terminal cleavage, resulting in an increase of MBP and intein-CBD species. On the other hand, the splicing activity of the T138S mutant was comparable with that of the wild type construct, and its N-terminal cleavage activity was less than that of the T138A mutant. The identity of the splicing and cleavage products was further confirmed by Western blot analysis with antibodies against MBP, the intein, and CBD (see supplementary Fig. 1 in the on-line version of this article). This observation suggests that replacement of Thr-138 with alanine does not affect the initial acyl shift; the inhibition could be because of the effect of the mutation on the transesterification or branch resolution step.

Conserved Catalytic Modules of Inteins—We further explored whether the catalytic site identified in this investigation is conserved among the inteins from diverse organisms. Both Asp-136 and His-143 are highly conserved residues in block F (Fig. 5*A*). Of the 136 inteins present in the data base (the New England Biolabs Intein Data Base) at the time of our comparison, 107 contain an aspartate residue in the corresponding position in block F. With only two exceptions, the other 27 inteins possess a serine, cysteine, threonine, or asparagine residue, which may play a role in hydrogen bonding to stabilize an oxyanion. In addition, 90 inteins possess a histidine residue at the position corresponding to His-143. A threonine residue is present in the block F position of 47 intein sequences, corresponding to Thr-138 of the Ssp DnaB intein, and a significant number of inteins possess serine, glutamate, or histidine (Fig. 5*A*). These residues may perform a similar role as Thr-138 in the Ssp DnaB intein in hydrogen bonding with the nucleophilic side chain at the C-terminal splice junction. In this study the conserved substitution (T138S) showed significantly higher splicing activity than the T138A mutant. The high level of amino acid conservation at these active site positions and their substitution with similar amino acids suggests their conserved roles in the splicing reaction. The C-terminal catalytic site structures of other inteins modeled on the Ssp DnaB intein can account for variations in blocks F and G. The crucial hydrogen-bonding interactions are retained in these inteins, suggesting a similar catalytic mechanism. Although more than 80% of the inteins retain a penultimate histidine, others primarily possess serine, glycine, or lysine (Fig. 5*A*). Serine or lysine may substitute for histidine in hydrogen bonding to an oxyanion. The *Methanococcus jannaschii* phosphoenolpyruvate synthase intein, representing those inteins lacking a penultimate histidine, could undergo splicing due to the presence of the Asn in block F. Indeed the Ssp DnaE intein, which lacks histidine in the penultimate position but has an aspartate in block F, is splicing-proficient (42–44).

This investigation suggests that the Ssp DnaB intein possesses two catalytic sites. The question of whether the chemical steps are catalyzed by a single or two separate active sites is a matter of debate; however, the catalytic sites responsible for N- and C-terminal cleavage steps in the Ssp DnaB intein do not appear to share common components and can proceed independently. The C-terminal cleavage reaction of the Ssp DnaB mini-intein, the *Pyrococcus* species B-D DNA polymerase intein, and the *Mycobacterium tuberculosis* RecA intein appear not to depend on the first two steps in the splicing pathway (12, 19, 45). On the other hand, because of the close proximity of the two junctions or active sites and the intricate balance of the three chemical steps, certain mutations introduced to one junction have been shown to affect the cleavage reaction at the other junction (14, 20). For instance, mutation of the penultimate histidine to glutamine in the Sce VMA intein resulted in a mutant in which C-terminal cleavage appeared to depend on thiol-inducible N-terminal cleavage (46). Because the hedgehog autoprocessing domains lack the essential intein block F and G elements (11), it is likely that inteins may have evolved as a new family of genetic elements by acquiring another catalytic site in their C-terminal subdomain. The answer may rely on the identification of autoprocessing proteins that possess activities analogous to the succinimide-based mechanism found in inteins. The understanding of inteins and their autocatalytic process have already led to numerous applications of inteins in protein science such as protein isolation, ligation, cyclization, screening for protein interactions, and containment of transgenes (47, 48). Hence there will be continued interest in inteins because of their emergence as one of the new paradigms for the understanding of catalysis and of structure-function relationships.

Acknowledgments—We thank Dr. Donald Comb and New England Biolabs, Inc. for generous support. We also thank Drs. Luet Wong, Thomas C. Evans, Chris Noren, William Jack, Nicole Nichols, and Prof. J. F. G. Vliegenthart for valuable discussion and reading of the manuscript.

REFERENCES

1. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M., and Stevens, T. H. (1990) *Science* **250,** 651–657
2. Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1990) *J. Biol. Chem.* **265,** 6726–6733
3. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J., and Belfort, M. (1994) *Nucleic Acids Res.* **22,** 1125–1127
4. Paulus, H. (2000) *Annu. Rev. Biochem.* **69,** 447–496
5. Xu, M. Q., Southworth, M. W., Mersha, F. B., Hornstra, L. J., and Perler, F. B. (1993) *Cell* **75,** 1371–1377
6. Perler, F. B. (2000) *Nucleic Acids Res.* **28,** 344–345
7. Chong, S., and Xu, M. Q. (1997) *J. Biol. Chem.* **272,** 15587–15590
8. Derbyshire, V., Wood, D. W., Wu, W., Dansereau, J. T., Dalgaard, J. Z., and Belfort, M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11466–11471
9. Shingledecker, K., Jiang, S. Q., and Paulus, H. (1998) *Gene* (*Amst.*) **207,** 187–195
10. Pietrokovski, S. (1998) *Protein Sci.* **7,** 64–71
11. Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) *Cell* **91,** 85–97
12. Xu, M. Q., and Perler, F. B. (1996) *EMBO J.* **15,** 5146–5153
13. Xu, M. Q., Comb, D. G., Paulus, H., Noren, C. J., Shao, Y., and Perler, F. B. (1994) *EMBO J.* **13,** 5517–5522
14. Hirata, R., and Anraku, Y. (1992) *Biochem. Biophys. Res. Commun.* **188,** 40–47
15. Cooper, A. A., Chen, Y. J., Lindorfer, M. A., and Stevens, T. H. (1993) *EMBO J.* **12,** 2575–2583
16. Davis, E. O., Jenner, P. J., Brooks, P. C., Colston, M. J., and Sedgwick, S. G. (1992) *Cell* **71,** 201–210
17. Shao, Y., Xu, M. Q., and Paulus, H. (1996) *Biochemistry* **35,** 3810–3815
18. Shao, Y., Xu, M. Q., and Paulus, H. (1995) *Biochemistry* **34,** 10844–10850
19. Mathys, S., Evans, T. C., Chute, I. C., Wu, H., Chong, S., Benner, J., Liu, X. Q., and Xu, M. Q. (1999) *Gene* (*Amst.*) **231,** 1–13
20. Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B., and Xu, M. Q. (1996) *J. Biol. Chem.* **271,** 22159–22168
21. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) *Cell* **89,** 555–564
22. Poland, B. W., Xu, M. Q., and Quiocho, F. A. (2000) *J. Biol. Chem.* **275,** 16408–16413
23. Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y., and Satow, Y. (2002) *J. Mol. Biol.* **316,** 919–929
24. Hu, D., Crist, M., Duan, X., Quiocho, F. A., and Gimble, F. S. (2000) *J. Biol. Chem.* **275,** 2705–2712

25. Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K., and Morikawa, K. (2000) *J. Mol. Biol.* **300,** 889–901
26. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., Jr., and Sacchettini, J. C. (1998) *Nat. Struct. Biol.* **5,** 31–36
27. Wu, H., Xu, M. Q., and Liu, X. Q. (1998) *Biochim. Biophys. Acta* **1387,** 422–432
28. Southworth, M. W., Benner, J., and Perler, F. B. (2000) *EMBO J.* **19,** 5019–5026
29. Chen, L., Benner, J., and Perler, F. B. (2000) *J. Biol. Chem.* **275,** 20431–20435
30. Wang, S., and Liu, X. Q. (1997) *J. Biol. Chem.* **272,** 11869–11873
31. Ghosh, I., Sun, L., and Xu, M. Q. (2001) *J. Biol. Chem.* **276,** 24051–24058
32. Pietrokovski, S. (1996) *Trends Genet.* **12,** 287–288
33. Evans, T. C., Jr., and Xu, M. Q. (1999) *Biopolymers* **51,** 333–342
34. Xu, M. Q., and Evans, T. C., Jr. (2001) *Methods* **24,** 257–277
35. Chen, X., Xu, M. Q., Ding, Y., Ferrandon, S., and Rao, Z. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58,** 1201–1203
36. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 488–492
37. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382
38. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
39. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
40. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
41. Bruice, T. C., and Benkovic, S. J. (1966) *Bioorganic Mechanisms*, W.A. Benjamin, Inc., New York
42. Wu, H., Hu, Z., and Liu, X. Q. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9226–9231
43. Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C., and Benkovic, S. J. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13638–13643
44. Evans, T. C., Jr., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X. Q., and Xu, M. Q. (2000) *J. Biol. Chem.* **275,** 9091–9094
45. Wood, D. W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999) *Nat. Biotechnol.* **17,** 889–892
46. Chong, S., Williams, K. S., Wotkowicz, C., and Xu, M. Q. (1998) *J. Biol. Chem.* **273,** 10567–10577
47. Evans, T. C., Jr., and Xu, M. Q. (2002) *Chem. Rev.* **102,** 4869–4883
48. Muir, T. W. (2003) *Annu. Rev. Biochem.* **72,** 249–289
49. Esnouf, R. M. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 938–940
50. Merrit, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524