Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein–intein junction

Alessandra Romanelli*†, Alexander Shekhtman†‡, David Cowburn‡, and Tom W. Muir*§

*Laboratory of Synthetic Protein Chemistry, The Rockefeller University, 1230 York Avenue, New York, NY 10021; and ‡New York Structure Biology Center, New York, NY 10027

Edited by Rowena G. Matthews, University of Michigan, Ann Arbor, MI, and approved March 9, 2004 (received for review October 13, 2003)

Protein splicing is a posttranslational autocatalytic process in which an intervening sequence, termed an intein, is removed from a host protein, the extein. Although we have a reasonable picture of the basic chemical steps in protein splicing, our knowledge of how these are catalyzed and regulated is less well developed. In the current study, a combination of NMR spectroscopy and segmental isotopic labeling has been used to study the structure of an active protein splicing precursor, corresponding to an N-extein fusion of the *Mxe* GyrA intein. The $^{1}J_{NC'}$ coupling constant for the (−1) scissile peptide bond at the N-extein–intein junction was found to be ≈12 Hz, which indicates that this amide is highly polarized, perhaps because of nonplanarity. Additional mutagenesis and NMR studies indicate that conserved box B histidine residue is essential for catalysis of the first step of splicing and for maintaining the (−1) scissile bond in its unusual conformation. Overall, these studies support the ''ground-state destabilization'' model as part of the mechanism of catalysis.

Protein splicing is an autocatalytic process in which an internal intein domain excises itself from a host protein, the extein. The first intein was discovered in 1990 as a spacer within the *Saccharomyces cerevisiae* vacuolar ATPase gene (1, 2). Since then, >100 putative members of this protein domain family have been identified from all three branches of life (3). In addition, a growing number of proteins have been shown to contain autoprocessing domains that are orthologous to inteins at the structural and/or mechanistic levels (e.g., refs. 4–7). There are three broad categories of inteins: (*i*) Maxi-inteins, the largest class, contain a homing endonuclease domain inserted into the core intein sequence (Fig. 1*A*). This endonuclease domain is not required for splicing activity and can be deleted (8). (*ii*) Mini-inteins lack a homing endonuclease domain and are consequently much smaller in size. (*iii*) Trans-splicing inteins lack a peptide linkage between the N- and C-terminal halves of the domain. Thus, the intein is split in two and splicing only occurs on reconstitution of the fragments. All three classes of inteins are characterized by several conserved sequence motifs (blocks A, B, F, and G) containing the critical residues for catalyzing the splicing reaction (Fig. 1*A*). Indeed, with one exception (noted below), there are no sequence requirements in the extein in order for splicing to take place, and so intein activity is promiscuous with respect to the extein.

The mechanism of protein splicing has been studied extensively by using biochemical approaches (Fig. 1*A*) (9). The first step in the standard protein splicing mechanism involves an N → S (or N → O) acyl shift in which the N-extein unit is transferred to the sidechain SH or OH group of a Cys/Ser residue, located at the immediate N terminus of the intein. Next, the entire N-extein unit is transferred to a second conserved Cys/Ser/Thr residue at the intein–C-extein boundary (+1 position within the C-extein) in a transesterification step. The resulting branched intermediate is then resolved through a cyclization reaction involving a conserved asparagine residue at the C terminus of the intein. The intein is thus excised as a C-terminal succinimide derivative. In the final step, an amide bond is formed between the two exteins following an S → N (or O → N) acyl shift. This is known to be a spontaneous chemical rearrangement (10) and presumably does not require the intein.

Although we have a reasonable overview of the various steps in protein splicing, the mechanistic details of autocatalysis remain poorly understood. There have been several high-resolution crystal structures of inteins (11–16) and related proteins (5, 7, 17). These structures have shed some light on the mechanism of the first step in protein splicing, the N → S (or N → O) acyl shift. As illustrated in Fig. 1*B*, the highly conserved block B His and Thr residues are located close to the scissile (−1) amide bond in all of the intein structures determined to date. The imidazole side chain of the conserved histidine (His-75 in *Mycobacterium xenopi* DNA gyrase A, *Mxe* GyrA) seems to be positioned to participate in general acid/base type catalysis, whereas the threonine appears to be part of a oxyanion hole that would stabilize the presumed tetrahedral oxythiazolidine intermediate (Fig. 1*B*).

Perhaps the most remarkable feature of these structural studies relates to the conformation of the scissile (−1) peptide bond itself. In cases where N-extein residues were actually present in the crystallized constructs, the (−1) amide has been found in a variety of conformations ranging from normal trans (15, 16) to distorted trans (13) to a cis configuration (12). This has led to the hypothesis that the scissile (−1) peptide bond is activated toward nucleophilic attack by distortion away from the most stable trans configuration, the strain being imposed by specific tertiary interactions involving conserved intein residues (Fig. 1*B*). Although intriguing, there are two major caveats to this "ground-state destablization" model; first, an unusual (−1) amide conformation is not found in all N-extein–intein crystal structures (15, 16), and second, all of the structures solved to date involved inteins inactivated through mutation of conserved residues. For example, in the *Mxe* GyrA intein structure, Cys-1 was mutated to a serine. Moreover, the construct contained only a single N-extein residue and consequently the positively charged α-amino group was very close to the active site of the intein. Conceivably, these mutations/alterations could have affected the (−1) amide conformation.

Clearly, it would be desirable to obtain a crystal structure of an active N-extein–intein precursor. This has proven difficult because of the short half-life of splicing precursors, on the order of a few hours (18, 19). Indeed, an active *S. cerevisiae* vacuolar ATPase (*Sce* VMA) intein has been shown to catalyze protein splicing within the crystalline lattice (15). Thus, there is no information currently available regarding the conformation of the scissile (−1) amide bond in an active N-extein–intein precursor. In the present study, we set out to address this issue by using a combination of solution NMR

BIOCHEMISTRY

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: *Sce* VMA, *Saccharomyces cerevisiae* vacuolar ATPase; RP-HPLC, reverse-phase HPLC; Gdm·Cl, guanidinium chloride; HSQC, heteronuclear single quantum correlation; *Mxe* GyrA, *Mycobacterium xenopi* DNA gyrase A.

†A.R. and A.S. contributed equally to this work.

§To whom correspondence should be addressed. E-mail: muirt@rockefeller.edu.

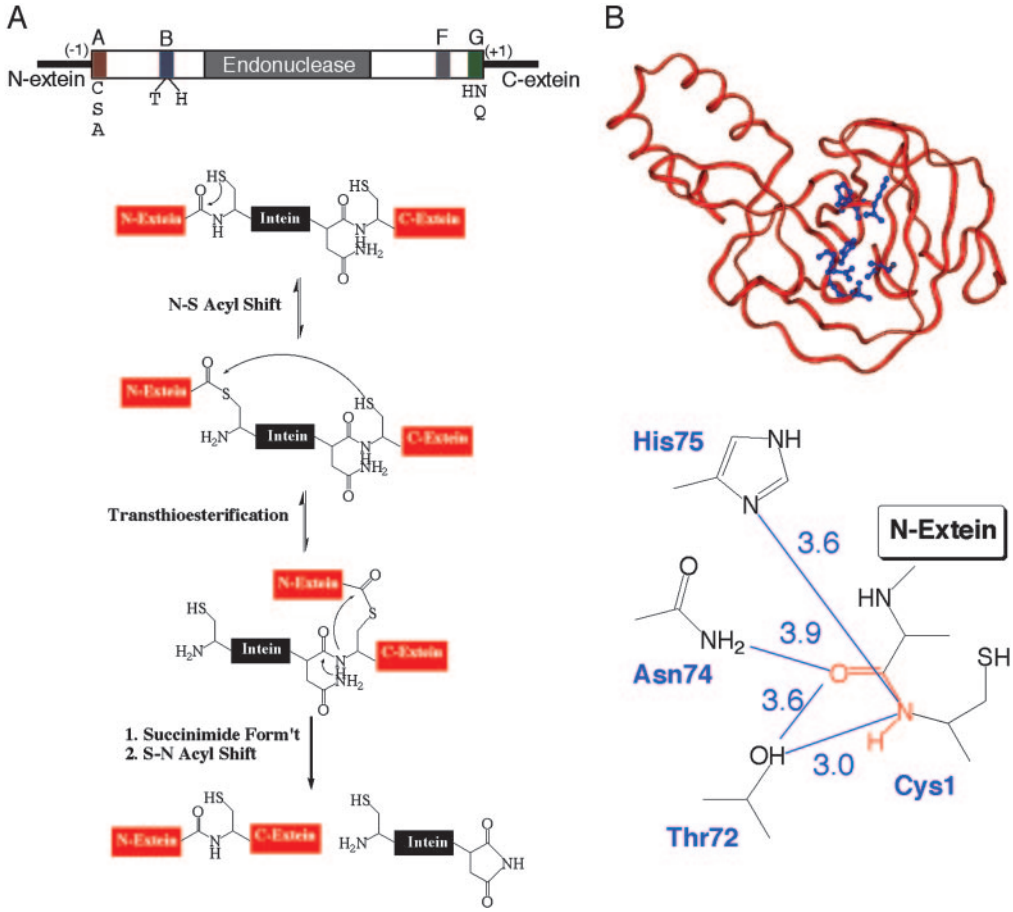

Fig. 1. The mechanism of protein splicing. (*A*) (*Upper*) Schematic illustrating the conserved regions within the intein domain family. The white bar represents the intein primary structure with conserved blocks A, B, F, and G indicated by colored boxes. The homing endonuclease domain found in the majority of inteins is also indicated. Conserved residues implicated in the splicing mechanism are shown below the bar. (*Lower*) The basic protein splicing mechanism. Note that a small number of inteins contain an Ala as the first residue, rather than Cys or Ser. These inteins are thought to directly use the +1 nucleophile within the C-extein in the first step (40). (*B*) (*Upper*) Ribbon representation of the *Mxe* GyrA crystal structure with the active site residues indicated in ball and stick (12). (*Lower*) Diagram of the N-terminal splice junction of the *Mxe* GyrA intein. The scissile (−1) amide is highlighted in red, and purple lines indicate possible hydrogen bond interactions to this amide with corresponding distances in Å.

spectroscopy and segmental isotopic labeling. Specifically, NMR was used to study the structure of the (−1) amide in an active protein splicing precursor, prepared by semisynthesis, corresponding to an N-extein fusion of the *Mxe* GyrA intein.

Materials and Methods

Peptide Synthesis. The peptide, H-AAMR[^{13}C′]F-SR, was synthesized on a 3-mercaptopropionamide-4-methylbenzhydrylamine (MBHA) resin (20) by using the *in situ* neutralization/2-(hydroxybenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) activation protocol for Boc-SPPS (21). After chain assembly, the peptide α-thioester was cleaved off the resin with anhydrous HF containing 4% vol/vol *p*-cresol (1 h, 4°C) and purified by preparative reverse phase HPLC (RP-HPLC) (Vydac C18, 15–20 μM, 50 mm × 250 mm) using a linear gradient of 13.5–31.5% B (9:1 MeCN/water, 0.1% trifluroacetic acid) over 60 min at a flow rate of 30 ml/min to give 40 mg of purified peptide [electrospray mass spectrometry (ESMS), obtained = 684.1 m/z $(M+H)^+$, expected = 682.9 m/z $(M+H)^+$].

Cloning and Protein Expression. The DNA encoding the *Mxe* GyrA intein (residues 1–198) was amplified by PCR using the pTXB1 vector (New England Biolabs) as a template. Note that pTXB1 encodes the GyrA intein with a single mutation, Asn-198 → Ala, which prevents cleavage of the intein–C-extein peptide bond. Importantly, this mutation does not affect the N-terminal splicing reaction (22), which is the focus of this study, and for clarity is referred to as the wild-type sequence hereafter. For construction of the plasmid *pTrcHis-Xa-GyrA*$_{wt}$, which encodes a factor Xa cleavage sequence between a poly(His) tag and the intein, the appropriate PCR product was cloned into the pTrcHisA vector (Invitrogen) by using *Bam*HI and *Hin*dIII restriction sites. This expression plasmid was used as a template for site-directed mutagenesis with the QuikChange site-directed mutagenesis kit (Stratagene). This resulted in the generation of the following expression vectors: *pTrcHis-Xa-GyrA*$_{T72V}$; *pTrcHis-Xa-GyrA*$_{N74A}$; *pTrcHis-Xa-GyrA*$_{H75A}$; *pTrcHis-Xa-GyrA*$_{T72VN74A}$; *pTrcHis-Xa-GyrA*$_{T72VH75A}$; and *pTrcHis-Xa-GyrA*$_{N74AH75A}$. The plasmid *pTrcHis-SH3-GyrA*$_{WT}$, which encodes the central Src homology 3 (SH3) domain of murine c-Crk-II between a poly(His) tag and the intein, was a gift from Vasant Muralidharan (Rockefeller University, New York). For protein expression, *Escherichia coli* BL21 (DE3) cells, transformed with the appropriate plasmid, were grown to mid-log phase at 37°C in LB medium. For ^{15}N-labeled proteins, the LB media was replaced with M9 minimal medium, supplemented with 0.2% (wt/vol) [^{15}N]NH_4Cl. Protein expression was induced with 0.4 mM isopropyl β-D-thiogalactoside (IPTG) at 37°C for 5 h, after which cells were harvested and lysed by passage through a French press. Fusion proteins were purified by affinity chromatography using a Ni^{2+} high-trap column (Amersham Pharmacia) and dialyzed into

Table 1. First-order rate constants for thiolysis of N-extein–GyrA intein fusions

Protein	K_{obs}, sec^{-1}	$t_{1/2}$
WT	$9.8 (\pm 0.9) \times 10^{-4}$	11.8 min
T72V	$2.5 (\pm 0.4) \times 10^{-4}$	46.2 min
N74A	$9.3 (\pm 0.8) \times 10^{-5}$	2 h
H75A	Trace	–
T72VN74A	$6.2 (\pm 0.3) \times 10^{-5}$	3.1 h
T72VH75A	Trace	–
N74AH75A	Trace	–

50 mM Tris, pH 8/1 mM EDTA buffer. All proteins were judged to be >90% pure based on SDS/PAGE, RP-HPLC, and ESMS analyses.

N-Terminal Cleavage Assay. Cleavage reactions were initiated by diluting (1:1) the freshly dialyzed His–Xa–GyrA fusion protein into DTT cleavage buffer (50 mM Tris, pH 8/100 mM DTT). Aliquots of the cleavage reactions were removed, quenched by adding guanidinium chloride (Gdm·Cl) (to a final concentration of 6 M), and then analyzed by analytical RP-HPLC. Reaction progress was monitored by integration of the starting material peak in the chromatograms. The data from the time courses were then fit to a first-order rate function (Table 1). All kinetic measurements were carried out at least in duplicate.

Generation of [U-^{15}N]GyrA and [U-^{15}N]GyrA(H75A). [U-^{15}N]GyrA was prepared by incubating the purified [U-^{15}N]His-SH3–GyrA fusion protein overnight in DTT cleavage buffer. [U-^{15}N]GyrA(H75A) was generated by incubating purified [U-^{15}N]His–Xa–GyrA(H75A) with factor Xa in proteolysis buffer (50 mM Tris, pH 8/0.1 M NaCl/1 mM $CaCl_2$) for 10 h at room temperature. Both proteins were then purified to >95% homogeneity by preparative RP-HPLC and characterized by ESMS: [U-^{15}N]GyrA, observed = 21,524 ± 3 Da, calculated = 21,529 Da; [U-^{15}N]GyrA(H75A), observed = 21,464 ± 4 Da, expected = 21,460 Da.

Preparation of AAMR[^{13}C′]F-[U-^{15}N]GyrA (1) and AAMR[^{13}C′]F-[U-^{15}N]GyrA(H75A) (2). Segmentally labeled proteins were obtained by chemically ligating H-AAMR[^{13}C′]F-SR with either [U-^{15}N]GyrA or [U-^{15}N]GyrA(H75A). Ligation reactions were initiated by dissolving purified, lyophilized peptide (10 eq) and protein (1 eq) in ligation buffer (6 M Gdm·Cl/0.1 M NaP_i at pH 8) containing 3% MESNA and 2% ethanethiol. The final concentration of protein was ≈1 mM, and the final concentration of peptide was ≈10 mM. In both cases, the ligation reaction was found to be complete after 5 h incubation at room temperature. The ligation products (**1** and **2**) were then purified by preparative RP-HPLC and characterized by ESMS: **1**, observed = 22,101 ± 4 Da, expected = 22,107 Da; **2**, observed = 22,045 ± 3 Da, expected = 22,038 Da. Purified proteins **1** and **2** (1 mg/ml) were folded by stepwise dialysis from denaturing buffer (6 M Gdm·Cl/0.1 M NaCl/20 mM KP_i, pH 6.6/1 mM DTT) into NMR sample buffer (0.1 M NaCl/20 mM KP_i, pH 6.6) at 4°C.

NMR Spectroscopy. NMR samples were prepared by concentrating (100 μM) labeled proteins (**1** and **2**) in NMR sample buffer containing 10% D_2O and 0.01% NaN_3. ^{1}H{^{15}N} HSQC (heteronuclear single quantum correlation) and 2D planes of HNCO spectra were collected at 4°C on a Bruker Avance-800 spectrometer with a cryo-probe. For the HSQC experiments, we collected 512 complex points in the ^{1}H and ^{15}N dimensions. In the 2D H{N}CO experiments, we collected 512 complex points in the ^{1}H dimension and 40 complex points in the ^{13}C dimension. Data sets were multiplied by a cosine-bell window function and zero-filled to 1,000 points before Fourier transformation using XWINNMR (Bruker Instruments). The corresponding sweep widths were 12.5 ppm, 12 ppm, and 30 ppm in the ^{1}H, ^{13}C, and ^{15}N dimensions, respectively. Experimental amide $^1J_{NC'}$ coupling constants were obtained by fitting the time evolution of the normalized peak intensities extracted from the series of HNCO-type experiments using the expression

$$I_k = e(-4t_1R_{2k}) \sin^2(2\pi J_k t_1), \quad [1]$$

where I_k are the normalized peak intensities for peak k, R_2 is the transverse relaxation time for peak k, and t_1 is the indirect dimension delay (23).

Results

The *Mxe* GyrA intein was chosen for our studies for several reasons. First, a high-resolution crystal structure of an inactive mutant of the protein is available (12), which was used to guide the mutagenesis studies described below. Second, this intein can be efficiently refolded from the denatured state (24), a property that was key to the semisynthesis of an active N-extein–intein precursor. Last, the relatively small size of this miniintein makes it more suitable for NMR studies than maxiinteins such as the *Sce* VMA intein.

Site-Directed Mutagenesis of the *Mxe* GyrA Intein. The crystal structure of the *Mxe* GyrA intein (12) revealed three residues in close proximity to the scissile (−1) peptide bond; Thr-72, Asn-74, and His-75, all from conserved block B (Fig. 1*A*). These residues are thought to be critical for activating the amide bond toward nucleophilic attack by the Cys thiolate. Mutation of the block B Thr and His residues has been found to abolish, or significantly impair, splicing in the *Sce* VMA intein (25); however, the effect of such mutations on *Mxe* GyrA intein activity has not been reported. Consequently, a series of N-extein–GyrA intein fusion proteins were generated in which the three block B residues were mutated individually or in pairwise combinations. These constructs contained an N-terminal His-tag, for purification, fused via a short peptide to the intein.

The ability of the intein mutants to catalyze the initial N → S acyl shift was determined by measuring the rate of thiol-induced cleavage of the N-extein peptide from the intein. The results of these experiments are summarized in Table 1. For the wild-type intein, the apparent first order rate constant for cleavage was $\approx 1 \times 10^{-3}$ s^{-1}, which is in excellent agreement with similar measurements made on other inteins (18, 19). Mutation of Thr-72 and Asn-74 reduced the cleavage rate by 5- and 10-fold, respectively, indicating that these residues are important, but not essential for splicing activity. Indeed, the double mutant Thr-72 → Val/Asn-74 → Ala was still able support robust N-terminal cleavage. In contrast, the His-75 → Ala mutation essentially abolished N-terminal splicing activity, and only trace amounts of cleavage products could be detected after 24 h of incubation in cleavage buffer. This result is consistent with His-75 having an essential catalytic role in the splicing activity of the GyrA intein; however, it is equally possible that the Ala mutation resulted in a gross reorganization of the intein structure. This possibility was investigated by NMR spectroscopy, as described below.

Semisynthesis of Segmental Isotopic-Labeled Proteins. Two segmental isotopic labeled constructs were prepared for NMR studies. These corresponded to a short N-extein peptide followed by either the wild-type GyrA intein sequence or the His75Ala mutant thereof. The N-extein sequence, AAMRF, was based on the native residues in the *gyrA* gene with the exception that, for synthetic reasons, the phenylalanine replaced a tyrosine. In both cases, the intein was uniformly labeled with ^{15}N, whereas the N-extein component contained a single ^{13}C isotope at the C′ position of the Phe.

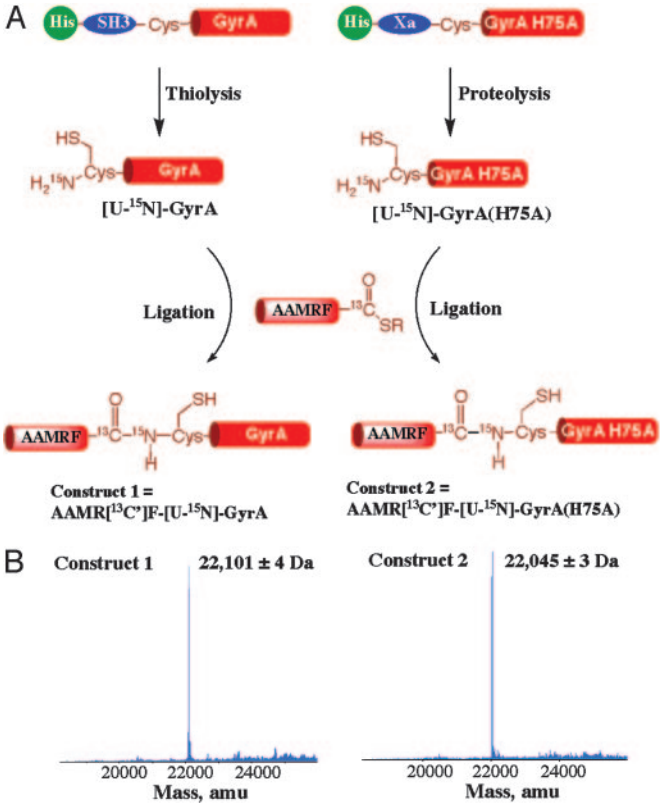

Fig. 2. Segmental isotopic labeling of N-extein–intein constructs. (*A*) Expressed protein ligation strategy used to prepare constructs 1 and 2. This semisynthetic route results in only the scissile (−1) amide being dual labeled with ^{13}C and ^{15}N. (*B*) Electrospray mass spectra of purified 1 and 2. Expected masses for 1 and 2 are 22,107 Da and 22,038 Da, respectively.

This segmental labeling pattern served two purposes; first, it allowed the folded state of the intein to be evaluated from the $^{1}H\{^{15}N\}$HSQC fingerprint, and second, it provided a unique spectral handle on the scissile (−1) peptide bond because only this amide is dual-labeled (Fig. 2*A*).

The two constructs AAMR[$^{13}C'$]F-[U-^{15}N]-GyrA (**1**) and AAMR[$^{13}C'$]F-[U-^{15}N]-GyrA(H75A) (**2**) were prepared by expressed protein ligation (EPL) (26). This involved chemical ligation of a synthetic peptide α-thioester, H-AAMR[$^{13}C'$]F-SR, to the appropriate uniformly ^{15}N labeled recombinant GyrA intein possessing an N-terminal Cys residue. The peptide building block was prepared by Boc-N^{a}-SPPS, whereas the protein building block was generated by thiolysis (wild-type GyrA) or proteolysis (mutant GyrA) of the appropriate GyrA-fusion protein (Fig. 2*A*). Ligation reactions were initiated by mixing an excess of the synthetic peptide with the intein in a buffer containing thiols and 6 M Gdm·Cl. The presence of denaturant was essential to prevent *in situ* thiolysis of the product during the ligation reaction; this would have occurred under native conditions, at least for the wild-type intein. Both ligation reactions were complete after ≈5 h as judged by HPLC and mass spectrometry (Fig. 2*B*). The ligation products, **1** or **2**, were each purified by preparative-scale RP-HPLC and then refolded by stepwise dialysis into nonreducing phosphate buffer at pH 6.5. Treatment of refolded construct **1** with N-terminal cleavage buffer resulted in quantitative cleavage of the free N-extein peptide from the intein, indicating that the protein had the native intein fold after the refolding procedure.

Because the principal goal of this study was to perform NMR experiments on active splicing precursor **1**, it was important to determine the stability of this molecule in NMR sample buffer (20 mM potassium phosphate, pH 6.6/0.1 M NaCl). This buffer does not contain any thiol reducing agents (unlike the DTT-containing buffers used in the N-cleavage reactions), and so N-terminal cleavage would have to involve hydrolysis of the thioester linkage rather than thiolysis, which should be quite a slow process at pH 6.6 and 4°C. Indeed, refolded protein **1** was found to be quite stable in this buffer, as determined by HPLC and MS analyses, with ≈10% N-terminal cleavage observed after 18-h incubation. Moreover, no evidence of cysteine oxidation was observed at this time point. As expected, the amount of cleavage products did accumulate with time, which places some restrictions of the duration of the NMR experiments to be performed. This problem is, to some extent, alleviated by the availability of cryogenic probes and ultra-high field NMR spectrometers.

Solution NMR Studies. NMR spectroscopy was used to study the solution structure of purified semisynthetic constructs **1** and **2**. Of initial interest was the effect of the His75Ala mutation on the global fold of the intein. Because constructs **1** and **2** each contain a uniformly ^{15}N-labeled intein domain, comparison of the $^{1}H\{^{15}N\}$ HSQC spectra should reveal any gross structural reorganization caused by the mutation. As shown in Fig. 3, there is in fact a high degree of spectral overlap between the two $^{1}H\{^{15}N\}$ HSQC fingerprint regions. Indeed, both spectra have well dispersed signals indicative of a defined tertiary fold with the expected predomi-

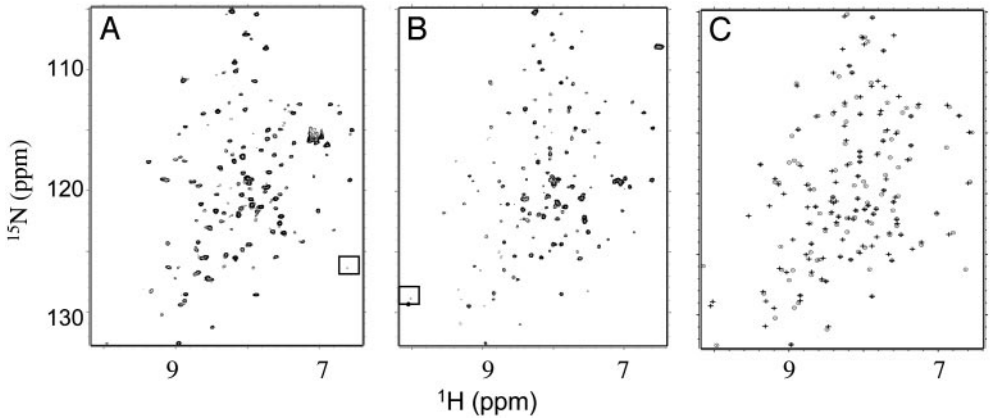

Fig. 3. Solution NMR spectra of segmental isotopic labeled constructs 1 and 2. (*A*) $^{1}H\{^{15}N\}$ HSQC spectra of AAMR[$^{13}C'$]F-[U-^{15}N]-GyrA (1). (*B*) $^{1}H\{^{15}N\}$ HSQC spectra of AAMR[$^{13}C'$]F-[U-^{15}N]-GyrA(H75A) (2). (*C*) Comparison of the reconstructed $^{1}H\{^{15}N\}$ HSQC spectra of constructs 1 (circles) and 2 (crosses) using chemical shifts extracted from individual spectra (*A* and *B*). Chemical shifts of Cys-1 HN bond are highlighted by squares.

nantly β-sheet structure (12). Although the vast majority of the resolvable crosspeaks are superimposable, a small number of signals can be observed that are unique to one or other of the spectra (highlighted in Fig. 3*C*). The spectral similarity factor (27) between the two is 16 Hz, which is indicative of some, but not dramatic, changes. Because the N-extein sequence is identical in proteins **1** and **2**, these differences presumably reflect localized structural changes and/or magnetic shielding effects resulting from removal of the imidazole ring. Nonetheless, it is clear that the His75Ala mutation, which renders the intein inactive, does not result in a global reorganization of tertiary structure. This argues that His-75 has an essential catalytic role in the splicing activity of the GyrA intein.

The dual-isotopic labeling pattern in constructs **1** and **2** allowed the unequivocal assignment of the scissile (−1) amide resonances from the HN plane of the HNCO spectra (Fig. 3 and supporting information, which is published on the PNAS web site). Comparison of these spectra reveals that the (−1) amide resonances of **2** are all shifted downfield relative to **1**. The most dramatic effect is on the amide proton, which shifts from 6.61 ppm in construct **1** to 10.01 ppm in construct **2**. The magnitude of this shift is remarkable and reflects the fact that the (−1) amide proton is unusually upfield shifted in the spectrum of **1** and downfield shifted in the spectrum of **2**.

Assignment of the scissile (−1) amide resonances allowed the one-bond dipolar coupling to be measured for constructs **1** and **2**. Experimental amide $^1J_{NC'}$ coupling constants were obtained by fitting the time evolution of the normalized peak intensities extracted from the series of HNCO-type experiments according to the method of Bax *et al.* (23). Accordingly, the $^1J_{NC'}$ coupling constants for the scissile (−1) amide in **1** and **2** were found to be 12.3 ± 0.3 Hz and 16.2 ± 0.2 Hz, respectively (Fig. 4).

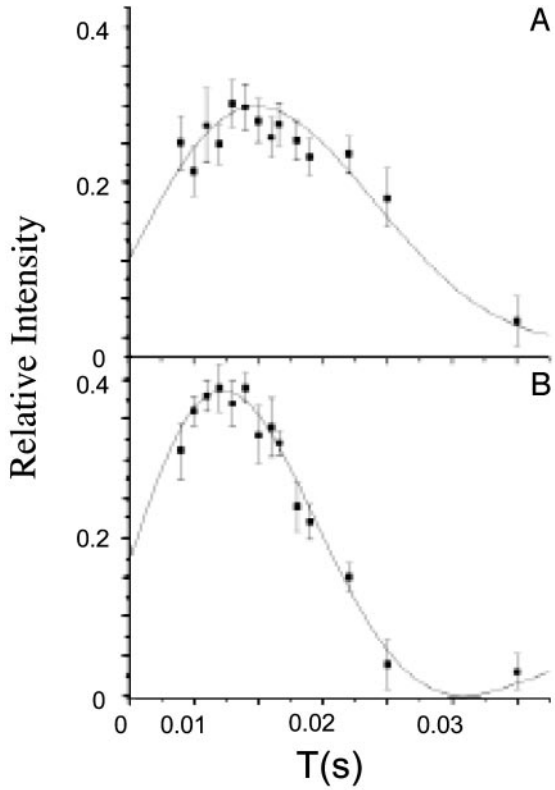

Fig. 4. Determination of the $^1J_{NC'}$ coupling constant for the scissile (−1) amide in constructs 1 and 2. The time evolution of the normalized peak intensities extracted from a series of H(N)CO experiments was nonlinearly fit to Eq. **1** to give the single-bond coupling constants. (*A*) Fit obtained for construct **1**, $^1J_{NC'}$ = 12.3 ± 0.3 Hz and R_2 = 17.3 ± 0.3 s^{-1}. (*B*) Fit obtained for construct **2**, $^1J_{NC'}$ = 16.2 ± 0.2 Hz and R_2 = 17.5 ± 0.4 s^{-1}.

Discussion

The first step in protein splicing involves an N → S (or N → O) acyl shift and results in the entire N-extein acyl unit being transferred to the side chain of the first residue of the intein. This rearrangement is chemically unfavorable because it breaks an amide bond and forms a high-energy (thio)ester linkage. With the exception of inteins and related autoprocessing proteins (e.g., refs. 4–7), this reaction is not known to occur spontaneously in proteins. Moreover, the reaction does not take place when an intein is chemically or thermally denatured. Thus, the chemical rearrangement must be catalyzed by structure–activity relationships within the intein. Several crystal structures of inteins are available. These reveal a homologous β-fold termed the HINT domain (also found in certain autoprocessing domains; ref. 17) and, not surprisingly, an active site sprinkled with conserved residues. However, these structures have provided less information about autocatalysis, in large part because they involve either free inteins or extein fusions of mutant inteins unable to catalyze splicing reactions. Thus, how the N → S(O) acyl shift is self-catalyzed by the intein is still an unsolved problem.

This study describes our initial efforts to understand the catalytic basis of protein splicing. Our view is that insights into autocatalysis will require the structural analysis of active protein splicing precursors involving, as far as possible, wild-type inteins and exteins. Such an approach is greatly complicated by the short half-life of extein–intein fusions (18, 19). Indeed, we have measured the half-life of an N-extein–GyrA fusion to be ≈12 min in the presence thiol cleaving agents (Table 1). To get around this problem, we used a combination of modern NMR instrumentation (i.e., ultra-high field spectrometers coupled with cryogenic probes) and site-specifically labeled protein splicing precursors. We devised a semi-synthetic route to N-extein fusions of the *Mxe* GyrA intein (Fig. 2) that allows site-specific incorporation of spectroscopic probes. In the present study, this was used to prepare multimilligram amounts of segmental isotopically labeled constructs, **1** and **2**, in which only the scissile (−1) amide bonds were dual-labeled with ^{13}C and ^{15}N isotopes. Importantly, this labeling pattern allowed the straightforward assignment of the (−1) amide resonances using 2D HNCO NMR experiments.

The chemical shift of the (−1) amide proton in **1** (i.e., NH of Cys-1) is significantly upfield shifted (δ = 6.61 ppm) compared to the random coil value for cysteine (28). In contrast, neither the ^{13}C nor ^{15}N chemical shifts in the (−1) amide are unusually upfield or downfield shifted (Fig. 3 and supporting information). Backbone amide ^1H chemical shifts are known to be sensitive to ring current fields and hydrogen-bonding networks (29). Inspection of the GyrA crystal structure (12) suggests that shielding effects from the nearest aromatic side chain, the imidazole ring of His-75, should be quite modest. Indeed, the chemical shift prediction program, SHIFTS (30), which uses the published coordinate file as input, predicts a small positive (0.01 ppm) ring current effect on the (−1) amide proton. Nonetheless, we do observe that removal of this imidazole ring, i.e., in construct **2**, results in a downfield shift of all of the (−1) amide resonances (Fig. 3) and so it is possible that the position of this side-chain is subtly different in construct **1** than in the GyrA intein construct used in the crystallographic study. It is also worth noting that constructs **1** and **2** both contain a phenylalanine at the first position in the N-extein. Conceivably, the ring current field from the phenyl ring could influence the (−1) amide chemical shifts in **1** and **2**. In terms of hydrogen bonding, short bond lengths are generally associated with downfield shifts of amide ^1H and longer bond lengths with upfield shifts (31). Based on the crystal structure, the side chains of residues Thr-72, Asn-74, and His-75 are all candidates for a putative long-range H-bonds to the (−1) amide (Fig. 1*B*).

BIOCHEMISTRY

Moreover, it is also possible that the scissile amide forms H-bonds to a residue(s) in the N-extein peptide. Indeed, in the crystal structure of glycosylasparaginase, which also undergoes an autoprocessing reaction involving an N → O shift, the scissile amide appears to be hydrogen-bonded to the side chain of an upstream Asp residue (5).

In addition to an unusual ^{1}H chemical shift, the scissile peptide bond in construct **1** has an abnormally low one bond dipolar coupling, $^1J_{NC'} = 12.3 \pm 0.3$ Hz. As with amide chemical shifts, amide $^1J_{NC'}$ values have been shown to correlate with hydrogen bonding, in this case H-bonding to the amide carbonyl increases the coupling constant, whereas H-bonding to the amide hydrogen decreases the coupling constant (32–34). This effect can be understood in terms of lengthening or shortening the C—N bond resulting from polarization of the π electrons across the amide. $^1J_{NC'}$ values have been reported for several proteins (34, 35) and are typically in the range 13–17 Hz. The measured $^1J_{NC'}$ coupling for the (−1) amide of **1** is among the lowest reported in a protein to date. This could indicate an extremely strong H-bond involving the NH group; however, this interpretation is inconsistent with the upfield chemical shift observed for this amide proton. Paradoxically, the $^1J_{NC'}$ coupling for the (−1) amide in inactive construct **2** is a more normal 16.2 ± 0.2 Hz, yet the amide ^{1}H exhibits a striking downfield shift compared to **1**, $\Delta\delta = 3.4$ ppm.

The backbone around the scissile (−1) peptide bond has been found to be in a distorted conformation in the crystal structures of two proteins, the *Sce* VMA intein (13) and glycosylasparaginase, which, as noted above, also undergoes autoproteolysis (5). Indeed, in the glycosylasparaginase structure, the planarity of scissile peptide bond (ω value) deviated from ideality ($\omega = 180°$) by >20°. Such a distortion would result in a significant loss in delocalization energy because of the breakdown π symmetry and it has been proposed that this activates the scissile peptide bond toward nucleophilic attack (5). Indeed, twisted amide bonds are known to be significantly more susceptible to alkaline hydrolysis than planar amides. Rate enhancements as high as 10^7 (10 kcal·mol^{-1}) have been reported (36), and recent theoretical studies suggest that the loss of π resonance is the major reason for this rate acceleration (37). The effect of C(O)—N bond rotation on NMR chemical shifts has been studied by using a series of twisted amides (38). A loss of amide planarity was associated with upfield and downfield shifts in the ^{15}N and ^{13}C amide signals, respectively. Based on these observations, one would expect that the amide ^{1}H chemical shift would also shift upfield as a result of C(O)—N bond rotation. The effect of amide nonplanarity on the $^1J_{NC'}$ coupling has not been studied in a systematic fashion. In principle, the $^1J_{NC'}$ coupling should decrease with increasing rotation of the C(O)—N bond because this will also increase the C(O)—N bond length (38). Indeed, simply density function theory quantum mechanical calculations on *N*-methylacetamide indicate that, as one might expect, $^1J_{NC'}$ varies significantly as the cosine of ω (supporting information). Surprisingly, Bax and coworkers (35) found no obvious correlation between the $^1J_{NC'}$ coupling and the nonplanarity of the peptide bonds in *Staphylococcal* nuclease. However, they did note that such a correlation could have been masked by uncertainties in the crystallographically determined ω values. It should also be added that the reported ω values for the backbone amides in this protein do not deviate significantly from ideality.

It is tempting to speculate that the extremely low $^1J_{NC'}$ coupling observed for the scissile amide in active construct **1** is evidence of backbone distortion, particularly because inactive construct **2** has a significantly higher $^1J_{NC'}$ coupling more typical of what is generally seen in proteins. The difference in $^1J_{NC'}$ between **1** and **2** is especially intriguing in light of two recent crystallographic studies on the *Sce* VMA intein (13, 15). One of these structures revealed backbone distortions around the scissile (−1) peptide bond, specifically τ for the −1 residue measured 100° (13), whereas these distortions were not observed in the other structure (15). Interestingly, the block B His was present in the construct used in the first structure but was mutated in the construct used in the second, just as in our constructs, **1** and **2**. Although additional studies, perhaps involving the use of multiple vicinal couplings (39), will be required to more precisely characterize the solution conformation of the scissile amide in the GyrA intein, our NMR data clearly indicate that the peptide bond in the GyrA intein is unusually polarized, and that the side chain of His-75 is absolutely required for this effect. These data provide additional support for the idea that the first step in protein splicing is facilitated, in part, by destabilizing the scissile amide bond. Future studies will need to explore this idea further and exploit our ability to prepare semisynthetic versions of the GyrA intein.

This work was supported by National Institutes of Health Grants GM55843 and EB001991 (to T.W.M.) and R01 GM47021 and P41-GM66254 (to D.C.). A.R. was supported by a Seconda Universitá di Napoli-Italy fellowship.

1. Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990) *J. Biol. Chem.* **265,** 6726–6733.
2. Kane, P. M., Yamashiro, C. T., Wolczyck, D. F., Neff, N., Goebl, M. & Stevens, T. H. (1990) *Science* **250,** 651–657.
3. Perler, F. B. (1999) *Nucleic Acids Res.* **27,** 346–347.
4. Porter, J. A., Ekker, S. C., Park, W.-J., von Kessler, D. P., Young, K. E., Chen, C.-H., Ma, Y., Woods, A. S., Cotter, R. J., Koonin, E. V. & Beachy, P. A. (1996) *Cell* **86,** 21–34.
5. Xu, Q., Buckley, D., Guan, C. & Guo, H.-C. (1999) *Cell* **98,** 651–661.
6. Rosenblum, J. S. & Blobel, G. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11370–11375.
7. Hodel, A. E., Hodel, M. R., Griffis, E. R., Hennig, K. A., Ratner, G. A., Xu, S. & Powers, M. A. (2002) *Mol. Cell* **10,** 347–358.
8. Chong, S. & Xu, M. Q. (1997) *J. Biol. Chem.* **272,** 15587–15590.
9. Giriat, I., Muir, T. W. & Perler, F. B. (2001) *Genet. Engineering* **23,** 171–199.
10. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. (1994) *Science* **266,** 776–779.
11. Duan, X., Gimble, F. S. & Quiocho, F. A. (1997) *Cell* **89,** 555–564.
12. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., Jr., & Sacchettini, J. C. (1998) *Nat. Struct. Biol.* **5,** 31–36.
13. Poland, B. W., Xu, M.-Q. & Quiocho, F. A. (2000) *J. Biol. Chem.* **275,** 16408–16413.
14. Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. & Morikawa, K. (2000) *J. Mol. Biol.* **300,** 889–901.
15. Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y. & Satow, Y. (2002) *J. Mol. Biol.* **316,** 919–929.
16. Ding, Y., Xu, M.-Q., Ghost, I., Chen, X., Ferrandon, S., Lesage, G. & Rao, Z. (2003) *J. Biol. Chem.* **278,** 39133–39142
17. Tanaka Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A. & Leahy, D. J. (1997) *Cell* **91,** 85–97.
18. Chong, S., Williams, K. S., Wotkowitcz, C. & Xu, M.-Q. (1998) *J. Biol. Chem.* **273,** 10567–10577.
19. Martin, D. D., Xu, M. Q. & Evans, T. C., Jr. (2001) *Biochemistry* **40,** 1393–1402.
20. Camarero, J. A., Cotton, G. J., Adeva, A. & Muir, T. W. (1998) *J. Pep. Res.* **51,** 303–316.
21. Schnölzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S. B. H. (1992) *Int. J. Pept. Protein Res.* **40,** 180–193.
22. Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q. & Perler, F. B. (1999) *BioTechniques* **27,** 110–120.
23. Cornilescu, G., Hu, J.-S. & Bax, A. (1999) *J. Am. Chem. Soc.* **121,** 2949–2950.
24. Valiyaveetil, F. I., MacKinnon, R. & Muir, T. W. (2002) *J. Am. Chem. Soc.* **124,** 9113–9120.
25. Kawasaki, M., Nogami, S., Satow, Y., Ohya, Y. & Anraku, Y. (1997) *J. Biol. Chem.* **272,** 15668–15674.
26. Muir, T. W. (2003) *Annu. Rev. Biochem.* **72,** 249–289.
27. Camarero, J. A., Shekhtman, A., Campbell, E. A., Chlenov, M., Gruber, T. M., Bryant, D. A., Darst, S. A., Cowburn, D. & Muir, T. W. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 8536–8541.
28. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991) *J. Mol. Biol.* **222,** 311–333.
29. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
30. Xu, X. P. & Case, D. A. (2001) *J. Biomol. NMR* **21,** 321–333.
31. Wagner, G., Pardi, A. & Wuthrich, K. (1983) *J. Am. Chem. Soc.* **105,** 5948–5949.
32. Juranic, N., Ilich, P. K. & Macura, S. (1995) *J. Am. Chem. Soc.* **117,** 405–410.
33. Juranic, N., Likic, V. A., Prendergast, F. G. & Macura, S. (1996) *J. Am. Chem. Soc.* **118,** 7859–7860.
34. Juranic, N., Moncrieffe, M., Likic, V. A., Prendergast, F. G. & Macura, S. (2002) *J. Am. Chem. Soc.* **124,** 14221–14226.
35. Delaglio, F., Torchia, D. A. & Bax, A. (1991) *J. Biomol. NMR* **1,** 439–446.
36. Blackburn, G. M., Skaife, C. J. & Kay, I. T. (1980) *J. Chem. Res.*, 294–295.
37. Lopez, X., Mujika, J. I., Blackburn, G. M. & Karplus, M. (2003) *J. Phys. Chem. A* **107,** 2304–2314.
38. Yamada, S. (1996) *J. Org. Chem.* **61,** 941–946.
39. Hu, J.-S. & Bax, A. (1997) *J. Am. Chem. Soc.* **119,** 6360–6368.
40. Southworth, M. W., Benner, J. & Perler, F. B. (2000) *EMBO J.* **19,** 5019–5026.