JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
ARTICLE
pubs.acs.org/JACS

pK_a Coupling at the Intein Active Site: Implications for the Coordination Mechanism of Protein Splicing with a Conserved Aspartate

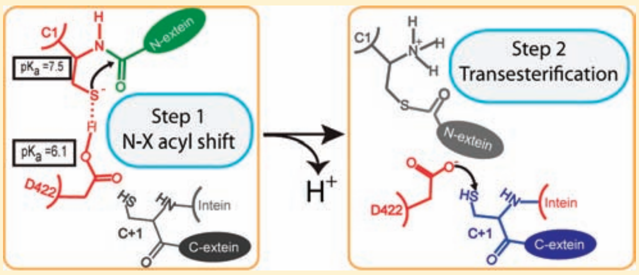
Zhenming Du,[†] Yuchuan Zheng,[‡] Melissa Patterson,[†] Yangzhong Liu,[‡] and Chunyu Wang*,[†]

[†]Biology Department, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

[‡]Department of Chemistry, University of Science and Technology of China, Hefei, Anhui, P.R. China 230026

S Supporting Information

ABSTRACT: Protein splicing is a robust multistep posttranslational process catalyzed by inteins. In the *Mtu* RecA intein, a conserved block-F aspartate (D422) coordinates different steps in protein splicing, but the precise mechanism is unclear. Solution NMR shows that D422 has a strikingly high pK_a of 6.1, two units above the normal pK_a of aspartate. The elevated pK_a of D422 is coupled to the depressed pK_a of another active-site residue, the block-A cysteine (C1). A C1A mutation lowers the D422 pK_a to normal, while a D422G mutation increases the C1 pK_a from 7.5 to 8.5. The pK_a coupling and NMR structure determination demonstrate that protonated D422 serves as a hydrogen bond donor to stabilize the C1 thiolate and promote the N−S acyl shift, the first step of protein splicing. Additionally, in vivo splicing assays with mutations of D422 to Glu, Cys, and Ser show that the deprotonated aspartate is essential for splicing, most likely by deprotonating and activating the downstream nucleophile in transesterification, the second step of protein splicing. We propose that the sequential protonation and deprotonation of the D422 side chain is the coordination mechanism for the first two steps of protein splicing.



INTRODUCTION

Protein splicing is a self-catalyzed post-translational process that results in the excision of an in-frame protein fusion, called an intein, from the precursor protein with the concomitant ligation of the two flanking polypeptides, called the N-extein and C-extein.[1] Since first reported in 1990,[2,3] protein splicing has found numerous applications in protein engineering, protein labeling, protein purification, and the control of protein function in vitro and in vivo.[4−7]

There are four steps of protein splicing in canonical inteins (Figure 1A): step 1, the N−X acyl shift (X = S or O); step 2, transesterification and the formation of a branched intermediate; step 3, asparagine cyclization coupled with C-terminal cleavage; and step 4, the X−N acyl shift and succinimide hydrolysis.[8,9] These steps are catalyzed by active site residues in conserved blocks in intein sequences (Figure 1B).[10] Block-A contains a conserved cysteine or serine at position 1, which provides a nucleophile for the N−X acyl shift. Block-B contains the highly conserved TxxH motif which catalyzes the N−X acyl shift.[11−13] Block-F has a conserved aspartate at position 4 (Figure 1B and C) and a conserved histidine at position 11. The conserved aspartate has been proposed to play a pivotal role in coordinating protein splicing, while the conserved histidine modulates C-terminal cleavage.[14−18] Block-G has a penultimate histidine at position 6 and a C-terminal asparagine at position 7, both critical for C-terminal cleavage. Block-G also includes the first residue of the C-extein, generally a Cys, Ser, or Thr and often called the +1 residue (e.g., C+1), and serves as the nucleophile for transesterification.

The multistep process of protein splicing is robust. Although off-pathway reactions, such as N- and C-terminal cleavage (Figure 1D), can be promoted by mutations,[16,18−20] they occur much less often than the on-pathway splicing reaction.[21−23] How are multistep reactions coordinated in protein splicing to ensure orderly progression of the splicing reaction and to minimize the side reactions? A number of general hypotheses exist in the literature, such as coordination by reaction rates, by mechanistic linkage, and by sequential conformational changes (reviewed by Mills and Paulus[24]). In a recent study of the *Mxe* GyrA intein, Frutos et al. have shown that transesterification and the formation of the branched intermediate (step two in protein splicing) speed up C-terminal cleavage (step three in splicing) by inducing a conformational change in the C-terminal splicing junction, thereby linking step two and step three by both reaction rate and conformational changes.[14] However, it is still unclear how step one (N−X acyl shift) and step two (transesterification) are catalytically linked in protein splicing. Without a mechanistic linkage, the reactive linear thioester formed by the N−X acyl shift

Received: April 7, 2011
Published: May 24, 2011

dx.doi.org/10.1021/ja203209f | *J. Am. Chem. Soc.* 2011, 133, 10275–10282

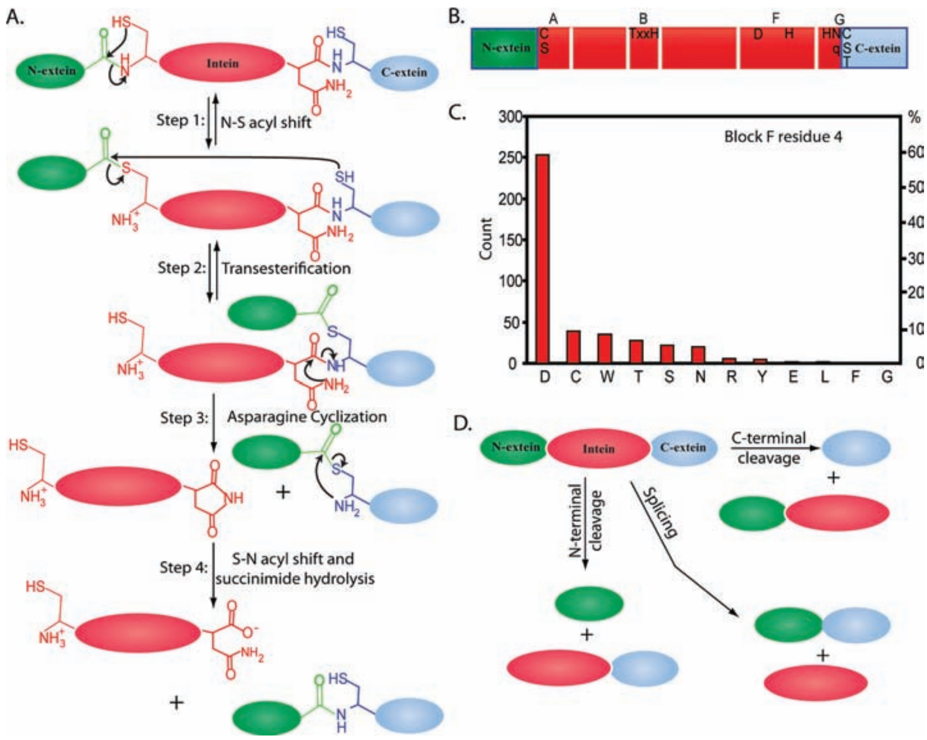


Figure 1. Key features of protein splicing. The N-extein, intein, and C-extein are consistently colored in green, red, and blue, respectively. (A) Four steps of protein splicing. (B) Conserved blocks in the intein sequence. Letters above define the blocks, and letters inside indicate conserved residues. (C) Residue frequencies at the block-F position 4. (D) Protein splicing and side reactions.

during step one could readily undergo premature N-terminal cleavage and compromise the efficiency of protein splicing.

The block-F aspartate D422 in the *Mycobacterium tuberculosis* (*Mtu*) RecA intein appears to be a prime candidate for coordinating protein splicing. D422 is conserved and present in 62% of all known intein sequences (Figure 1C). A D422G mutation, termed C-terminal cleavage mutant (CM), dramatically enhances the C-terminal cleavage reaction with minimal protein splicing, indicating the disruption of the coordination mechanism by this mutation.[16] In crystal structures, D422 is positioned adjacent to both N- and C-terminal splicing junctions and has multiple conformations in various engineered *Mtu* RecA inteins.[17] Recently, the Belfort and Nayak groups provided evidence for an essential role of D422 in the formation of a branched intermediate by activating the downstream nucleophile C+1.[18] Previously we have characterized the pK_a values and the mechanistic role of the most conserved block-B histidine and have studied the structure and dynamics of *Mtu* RecA intein mutants using solution NMR.[25−28] In this paper, we studied the ionization state, solution conformation, and the mechanistic role of the block-F aspartate, D422. We found that the pK_a values of D422 and C1 are coupled and that the intein active-site residues most likely form a network for specific protonation and deprotonation events during protein splicing. The protonated D422 facilitates the first step of protein splicing, by serving as a hydrogen bond donor to the C1 thiolate. We also show that D422C and D422S mutants have very low splicing activity, indicating that a protonated side chain alone is not sufficient for splicing. Deprotonated D422 likely activates C+1,[18] and therefore the sequential protonation (step 1) and deprotonation of the D422 side chain (step 2) will naturally link and coordinate the first two steps of protein splicing. Our results provide novel insights toward a complete coordination mechanism of protein splicing at the atomic resolution.

MATERIALS AND METHODS

Sample Preparation. NMR sample preparation of ΔΔIhh-SM and the inactive precursor (NI) was described previously.[26,27] For constructs 3−7 (Table 1), cells were grown at 37 °C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.5−0.6 and then incubated at 20 °C for 4 h. For isotopic labeling, cells were grown at 37 °C and induced with 1 mM IPTG at an OD_{600} of 0.3−0.4, and then the culture was grown for 20−24 h at 20 °C. Isotope labeling was accomplished by growing cultures in M9 minimal medium containing either 1 g/L of $^{15}NH_4Cl$ for uniformly ^{15}N-labeled samples or 1 g/L of $^{15}NH_4Cl$ and 1 g/L of $^{13}C_6$-D-glucose (Cambridge Isotope Laboratories, Andover, MA, USA) for uniformly ^{15}N- and ^{13}C-labeled samples, respectively. Cells were lysed by sonication in 20 mM Tris buffer, pH 8.0, and purified by ion exchange on Q-Sepharose (GE Healthcare, Uppsala, Sweden) followed by gel filtration chromatography on a HiLoad 16/60 Superdex 200 preparation grade column (GE Healthcare, Uppsala, Sweden). The pure protein was concentrated and exchanged by ultrafiltration with Amicon Ultra Centrifugal Filter Devices (Millipore Cooperation, Billerica, MA, USA) (10 kDa molecular weight cutoff) into 50 mM sodium phosphate buffer, pH 7.0, with 100 mM NaCl and 1 mM NaN_3. For NMR experiments, the final protein concentration varied between 0.15 and 1 mM with 10% D_2O for deuterium locking.

In Vivo Splicing Assay. JM109 cells transformed with pMIC express a fusion protein with maltose binding protein as the N-extein (M), mini-intein ΔIhh-SM (I), and a C-terminal domain of I-TevI as the C-extein (C).[29] The splicing assay was carried out as described previously.[18,27] Western blots of the cell lysate were carried out using

Table 1. Summary of Constructs Used in This Study

construct	name	N-extein	residue at C40 of CBD	first res. of intein	block-F position 4	usage in the study
1	ΔΔIhh-SM	---	---	C	D	NMR structure, D422 pK_a
2	NI (N-extein+ intein)	17 AA[a]	---	A	D	D422 pK_a with C1A
3	CBDΔΔIhh-SM	64 AA[b]	C	C	D	D422 pK_a w/N-extein
4	CBDΔΔIhh-SMD422E	64 AA[b]	C	C	E	D422 assignment in CBD-intein
5	CBDΔΔIhh-SMC40S	64 AA[b]	S	C	D	constructs 5−7
6	CBDΔΔIhh-SMC40SC1S	64 AA[b]	S	S	D	C1 pK_a by UV @ 240 nm which necessitates C1S and C40S mutants
7	CBDΔΔIhh-CMC40S	64 AA[b]	S	C	G	
8	pMIC-D422	MBP	---	C	D	
9	pMIC-D422C	MBP	---	C	C	
10	pMIC-D422S	MBP	---	C	S	
11	pMIC-D422T	MBP	---	C	T	constructs 8−15
12	pMIC-D422W	MBP	---	C	W	in vivo splicing assay
13	pMIC-D422E	MBP	---	C	E	
14	pMIC-D422GD24G	MBP	---	C	G	
15	pMIC-D422C1AN440A	MBP	---	A	D	

[a] N-extein = MAHHHHHHVGTGSNADP. [b] N-extein is the chitin binding domain (CBD).

the WesternBreeze kit (Invitrogen, Carlsbad, CA, USA), polyvinylidene fluoride membrane, and anti-C antibody.

Cysteine pK_a Determination. Cysteine pK_a was monitored using a NanoDrop UV spectrometer (ThermoScientific, Wilmington, DE, USA) with 180 μM protein. The extinction coefficient at 240 nm (ε_{240}) was calculated using $\varepsilon_{240} = \varepsilon_{280} \times A_{240}/A_{280}$. A_{240}/A_{280} is the ratio of the protein absorbance at 240 and 280 nm; ε_{280} is the extinction coefficient of the intein precursor at 280 nm (40 575 M^{-1} cm^{-1}); and ε_{240} is the extinction coefficient at 240 nm. All titrations were performed in triplicate, and average ε_{240} values were calculated. The data were fit to the Henderson−Hasselbalch equation

$$\varepsilon_{240}^{\mathrm{obs}}(\mathrm{pH}) = \varepsilon_{240}^{\mathrm{SH}} + \frac{\varepsilon_{240}^{\mathrm{S}^-} - \varepsilon_{240}^{\mathrm{SH}}}{1 + 10^{(\mathrm{p}K_a - \mathrm{pH})}}$$

where $\varepsilon_{240}^{\mathrm{obs}}(\mathrm{pH})$ is the observed extinction coefficient at 240 nm as a function of pH and $\varepsilon_{240}^{\mathrm{SH}}$ and $\varepsilon_{240}^{\mathrm{S}^-}$ are the extinction coefficients at 240 nm when cysteine is in thiol form and thiolate form, respectively.[30,31] Nonlinear regression was performed using Origin (Microcal Origin, Northampton, MA, USA).

Aspartate pK_a Determination. An NMR sample at pH 7.0 was split into two, one for titration from pH 7 to 10 and the other for titration from pH 7 to 3.5. The pH was adjusted with 0.1 M HCl and NaOH. For ΔΔIhh-SM, 2D HB(CB)CO spectra were acquired. For CBDΔΔIhh-SM, ^{13}C−^{1}H HSQC was used to monitor the chemical shift change of the side chains. Data were fit to the Henderson−Hasselbalch equation

$$\delta_{\mathrm{obs}} = \frac{\delta_{\mathrm{HA}} + \delta_{\mathrm{A}^-} 10^{(\mathrm{pH} - \mathrm{p}K_a)}}{(1 + \delta_{\mathrm{A}^-} 10^{(\mathrm{pH} - \mathrm{p}K_a)})}$$

where δ_{obs}, δ_{HA}, and δ_{A^-} are the chemical shifts for the observed, protonated, and deprotonated species, respectively. The pK_a values for ΔΔIhh-SM and NI were determined in 100% D_2O. The pK_a values were corrected for D_2O according to Krezel et al.[32]

NMR Resonance Assignment. For ΔΔIhh-SM and NI, the backbone resonances were assigned using HNCACB and HNCOCACB at 25 °C, and aspartate side chains were assigned using ^{15}N-TOCSY, ^{13}C−^{1}H HSQC, and HBCBCO spectra (Supporting Information, Figure S1). For CBDΔΔIhh-SM, D422 side chain resonances were assigned by the D422E mutant (Supporting Information, Figure S4). Other side chain assignments were carried out using standard NMR methods.

NMR Structure Determination. Distance restraints from 3D ^{15}N NOESY and ^{13}C NOESY were classified as strong (1.8−3.0 Å), medium (1.8−4.0 Å), weak (1.8−5.0 Å), and very weak (1.8−6.0 Å) based on cross-peak intensities. A total of 1473 distance restraints were used. H-bonds were identified by slow HD exchange rates and restrained to an upper limit of 3.5 Å between heteroatoms (N−H···O) and 2.4 Å between proton and the H-bond acceptor oxygen. Backbone dihedral angle restraints (φ and ψ) were derived from TALOS.[33] The initial 100 structures were calculated using Xplor-NIH 2.19 without RDC restraints.[34] D_a and rhombicity R for the alignment tensor were initially estimated using PALES.[35] The lowest energy conformation was selected for refinement with RDC values obtained from stretched gel[36] and d(GpG).[37] The statistics of the 20 lowest energy conformers out of 200 refined structures are summarized in Table 2. The quality of the structures was analyzed using the psvs server.[38]

RESULTS AND DISCUSSION

Elevated pK_a of D422 Highlights the Mechanistic Importance of the Protonated Aspartic Acid Side Chain. To explore the coordination mechanism of D422, we first examined the ionization state of D422 in an intein and in precursor fusion proteins with N-exteins. The ionization states of active site residues provide crucial information for their specific roles in acid−base catalysis.[27,30,39,40] Solution NMR is the most powerful experimental tool for site-specific pK_a determination in enzymes. We used 3D HBCBCO (Supporting Information, Figure S1)[41] to assign the aspartate side chain and fit ^{13}CO or H_β chemical shifts to the Henderson−Hasselbalch equation to determine the aspartate pK_a. Both ^{13}CO and H_β chemical shifts showed a pK_a of ~6 (Figure 2A and 2B) for D422 in a minimized *Mtu* RecA intein, ΔΔIhh-SM[16,29,42] (construct 1 in Table 1). This pK_a value is >2 units higher than the normal aspartate pK_a of ~3.9 and the pK_a values of the other aspartates in ΔΔIhh-SM (Supporting Information, Figure S2). The addition of an N-extein does not significantly perturb the pK_a of D422, as demonstrated in CBDΔΔIhh-SM (Figure 2B) (construct 3 in Table 1, pK_a ~5.8), which contains a chitin-binding domain (CBD) as the N-extein. Very few aspartates exhibit pK_a as high as 6,[43] and most of them play crucial catalytic roles in the protonated form, such as D25 in HIV-protease[44] and

Table 2. Statistics of NMR Structure Determination

(a)	NMR restraints	ΔΔIhh-SM
	distance restraints	1473
	intraresidue ($\|i-j\| = 0$)	304
	sequential ($\|i-j\| = 1$)	509
	medium range ($1 < \|i-j\| < 5$)	117
	long range ($\|i-j\| \geq 5$)	543
	dihedral angle restraints	162
	hydrogen bonds	81
	residual dipolar couplings	212
	(stretched gel and d(GpG))	
(b)	statistics for 20 lowest energy structures	
	rmsd from idealized covalent geometry	
	bonds (Å)	0.011
	angles (deg)	1.4
(c)	coordinate precision	bb/heavy
	rmsd to the mean (Å)	0.4/0.9
(d)	Ramachandran plot (% residues)	
	residues in most favored regions	89.2
	residues in additional allowed regions	9
	residues in generously allowed regions	1.8
	residues in disallowed regions	0.1

D61 in subunit C of ATP synthase.[45] Thus, the elevated pK_a of D422 strongly suggests that the protonated side chain of D422 plays an important role in the acid−base catalysis of protein splicing.

Protonated D422 Side Chain Serves As a Hydrogen Bond Donor for C1 Thiolate to Facilitate the N−S Acyl Shift. *pK_a Coupling between D422 and C1.* Both hydrophobic environment and local interactions may elevate the pK_a. D422 is 11% solvent exposed in an X-ray structure (PDB: 2IN0)[17] and our NMR structure (vida infra), indicating partial burial in the hydrophobic core. However, other aspartates with similar levels of solvent exposure, e.g., D91 (12%) and D97 (11%), have a normal pK_a value of ∼4 (Supporting Information, Figure S2). Therefore, the hydrophobic environment alone cannot account for the elevated pK_a of D422. Because C1 and D422 are spatially close and may interact with each other, we determined the pK_a of D422 in NI (a fusion protein of N-extein and intein), which contains a C1A mutation (construct 2 in Table 1). Surprisingly, the pK_a of D422 drops back to 4.0 (Figure 2A). This suggests that the D422 pK_a is raised by the highly conserved C1 side chain.

As C1 elevates the pK_a of D422, we tested if the D422 residue in turn affects the pK_a of the C1 thiol group, whose ionization status will critically affect the N−S acyl shift, the first step of protein splicing. We examined the C1 pK_a in the presence and absence of D422 in two intein precursors by measuring UV absorbance of the cysteine thiolate group at 240 nm[31] (constructs

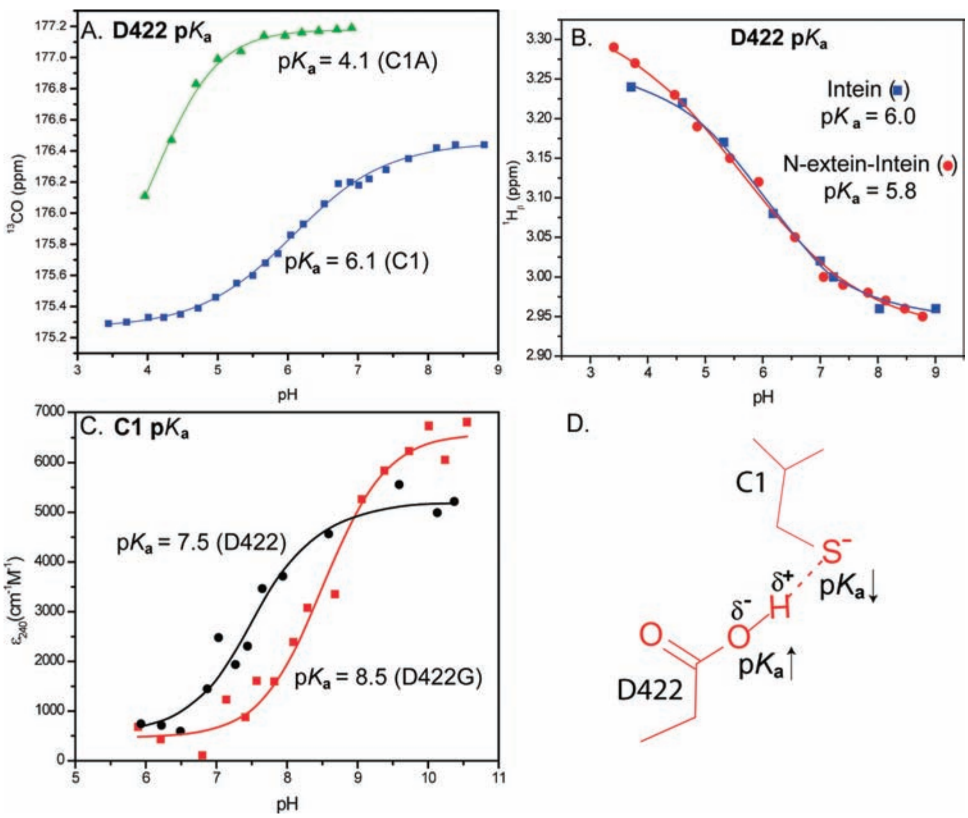


Figure 2. pK_a coupling between D422 and C1. (A) Block-F D422 pK_a is elevated to 6.1 in an engineered *Mtu* RecA intein (blue symbol, construct 1 in Table 1) as determined by ^{13}CO chemical shift titration. C1A mutation lowers D422 pK_a to normal (green symbol, construct 3 in Table 1). (B) The D422 pK_a is 5.8 in a precursor with an N-extein (red curve, construct 2 in Table 1), while D422 pK_a is 6.0 without N-extein (blue curve, construct 1 in Table 1), as determined by pH titration of H_β. (C) C1 has a depressed pK_a of 7.5 in the intein precursor with D422 (construct 5 in Table 1) and a pK_a of 8.5 with a D422G mutation (construct 7 in Table 1). (D) The structural basis for pK_a shift and coupling between C1 and D422, representing the catalytically active state.

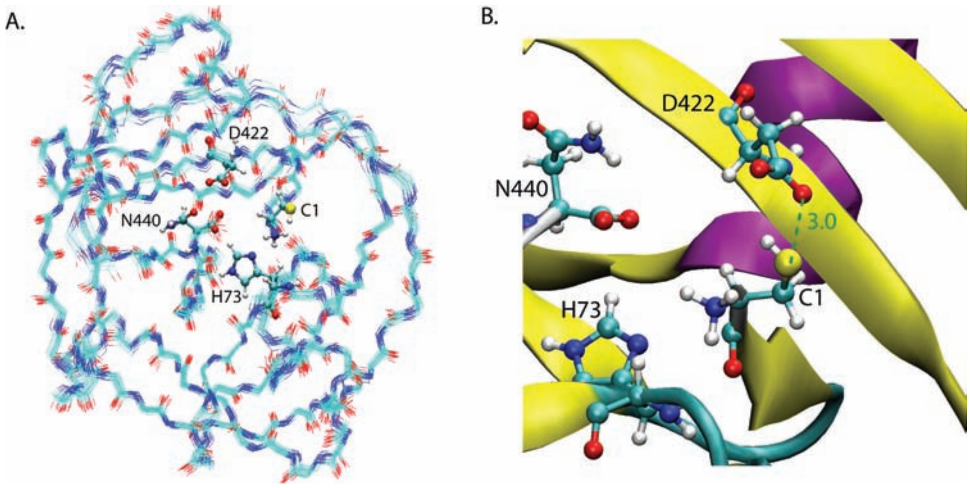


Figure 3. Solution NMR structures of ΔΔIhh-SM and the hydrogen bond between the block-F D422 side chain and C1. Active site residues C1, H73, D422, and N440 are shown in ball and stick representations. (A) Ensembles of 20 low-energy NMR structures determined for ΔΔIhh-SM, with a backbone rmsd of 0.4 Å. (B) D422 serves as a hydrogen bond donor to C1.

5 and 7 in Table 1). As reportedly earlier,[30] all interfering cysteines need to be mutated. We mutated the only other cysteine in constructs 5 and 7 (C40 in CBD) to serine (C40S). A construct with the additional C1S mutation (mutating C1 in intein to serine) was used as the baseline reference for UV absorbance at 240 nm (construct 6, Table 1). In the presence of D422, C1 has a depressed pK_a of 7.5, about one unit lower than the normal pK_a of cysteine, ~8.3 (Figure 2C). With a D422G mutation, in the absence of the aspartate side chain, C1 pK_a increases to 8.5 (Figure 2C). This suggests that the ~1 unit pK_a depression of C1 in the wild-type intein is due to the interactions between the C1 thiol group and the D422 side chain. This coupling between D422 pK_a elevation and C1 pK_a depression can be explained by a hydrogen bond model with the D422 hydroxyl serving as the donor and C1 thiolate serving as the acceptor (Figure 2D). For an ionizable group, a nearby positive charge lowers its pK_a, while a nearby negative charge increases its pK_a. Because D422 and C1 are close to each other in the 3D structure of the intein (Figure 3), the partial positive charge of the proton on the D422 hydroxyl group lowers the C1 pK_a, while the negative charge of the C1 thiolate raises the pK_a of D422 (Figure 2D). Such a hydrogen bond model explains the pK_a elevation/depression coupling of D422/C1 (Figure 2D, left panel) observed here. In contrast, a hydrogen bond model with the C1 thiol serving as the donor and D422 serving as the acceptor cannot explain the experimental pK_a values. In such a scenario, the partial positive charge of the thiol proton would have stabilized D422 in the deprotonated form, lowering its pK_a, while the deprotonated and negatively charged D422 would have increased the C1 pK_a, generating pK_a shifts opposite to those observed in this study.

Existence of the Proposed Hydrogen Bond between D422 and C1 in NMR Structures. In crystal structures of various engineered *Mtu* RecA inteins, D422 adopts multiple conformations.[17] To assess D422 conformation in solution, we determined the solution NMR structure of ΔΔIhh-SM to a backbone rmsd of 0.4 Å (Figure 3A; for structural statistics, see Table 2; PDB code: 2L8L), using NOE-derived distance constraints and residual dipolar couplings (RDC) measured in both stretched gel[36] and dGpG.[37] The NMR structure agrees well with the crystal structure (2IN0) with a CA rmsd of 1 Å. As expected, ΔΔIhh-SM has a horseshoe-like hedgehog-intein fold, with active site residues clustered in a cleft on the surface of the structure.[17] Of the 20 lowest energy NMR structures, four of them exhibit a hydrogen bond between D422 and C1 (Figure 3B), which further confirms the hydrogen-bond model (Figure 2D). Such a hydrogen bond can lower the activation energy of the N−S acyl shift by stabilizing the negatively charged thiolate of C1. A similar structural mechanism has been proposed in human DJ-1 protein where the pK_a depression of the active site cysteine is caused by a protonated glutamate serving as the hydrogen bond donor.[46] The importance of D422 in the N−S acyl shift has been demonstrated using mutagenesis and FRET by Belfort et al.[18] Among wild-type Asp and mutants Glu, Asn, Leu, Gly, and Ala at position 422, N-terminal cleavage proceeds with the highest rate for D422 and the next highest rate for E422.[18] We therefore conclude that protonated D422 promotes the N−S acyl shift, the first step of protein splicing, by serving as a hydrogen bond donor to C1, lowering its pK_a, and enhancing its reactivity. Because the N−S acyl shift can still proceed without an aspartate at the 422 position,[18] D422 may play a secondary role. We hypothesize that the highly conserved block-B H73 plays a primary role by directly activating the C1 thiol group, as proposed in our previous work.[27]

In Vivo Protein Splicing Assay Demonstrates Mechanistic Importance for the Deprotonated Aspartate Side Chain. To test whether protonated D422 is sufficient for splicing, we tested various mutants using a well-established in vivo splicing assay.[18,27,42] We replaced D422 with C, S, T, and W that occur with high frequency at position 422 in intein sequences (Figure 1C). The effect of D422 to C, S, T, and W mutations on protein splicing has not been previously tested. Importantly, C, S, T, and W have side chain groups that can serve as hydrogen bond donors to C1 in the same manner as D422. In the in vivo splicing assay, *E.coli* cells express an intein precursor MIC, composed of maltose binding protein (MBP) as the N-extein, the ΔI-SM intein, and the C-terminal domain of I-TevI as C-extein (constructs 8−15 in Table 1). The precursors may undergo splicing and/or N- and C-terminal cleavage (Figure 1D). Protein splicing generates bands corresponding to splice products, intein (I in Figure 4A), and ligated extein (MC in Figure 4A and B). The identities of the bands in

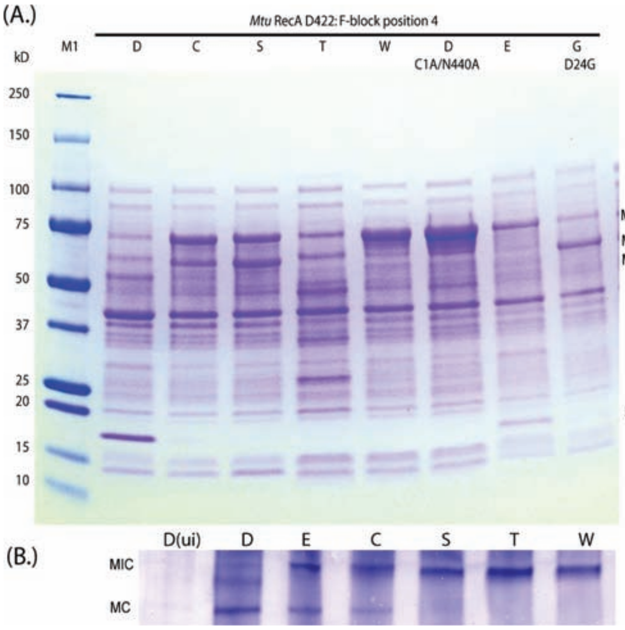


Figure 4. In vivo splicing assay demonstrates the importance of a deprotonated D422 side chain. (A) SDS-PAGE using MIC constructs with various mutations at the 422 position (constructs 8−15 in Table 1). Splicing precursor MIC, splice product MC, and excised intein I, and C-terminal cleavage product MI are labeled. (B) Western blot with an antibody for C-extein in the MIC construct.

SDS-PAGE were confirmed by Western blot,[18,27,42] as illustrated in Figure 4B where MIC and MC bands were visualized using an antibody against the C-extein.

With ΔI-SM as the intein (D422), I and MC were generated by protein splicing, and little precursor MIC remained (Figure 4A). The mutations with known phenotypes behaved as expected. For the negative control with double mutations of C1A and N440A, there was accumulation of the MIC precursor without any splicing or cleavage product. For the D442G mutation which promotes C-terminal cleavage,[16] a large amount of C-terminal cleavage product MI was observed. The D422E mutation resulted in a reduced amount of splice products I and MC and an increased amount of precursor, indicating E422 coordinates splicing with reduced efficiency, as shown by Belfort et al.[18] D422C mutation led to an even smaller amount of splice product I and MC, while D422S only generated a trace amount of I and MC (Figure 4A and B). For D422T and D422W, there was no splice product, indicating the disruption of the protein splicing mechanism. The bulkier side chains of threonine and tryptophan may interfere with the proper alignment of the intein active site.

These data demonstrate that a hydrogen bond donor at the D422 position is not sufficient for the efficient coordination of splicing. For side chains of similar size, the splicing efficiency is in this order: D > C > S (Figure 4A and B). Interestingly, the pK_a values of these amino acid side chains are inversely ordered: D < C < S. Thus, splicing efficiency appears inversely correlated to the side chain pK_a at the 422 position, with higher pK_a leading to lower splicing efficiency. Even though D422 has a higher than normal pK_a, too high a pK_a would be detrimental for splicing. This likely indicates that the deprotonated form of D422 also plays an essential role in coordinating splicing, in addition to the protonated form of D422.[18] Recently, it was shown that D422 is crucial in forming the branched intermediate in *Mtu* RecA intein and that in QM/MM simulation a proton from the C+1 thiol group is spontaneously transferred to the deprotonated D422 side chain.[18] Our data further support such a mechanistic role for deprotonated D422.

Coordination Mechanism of D422 Likely Involves Sequential Protonation and Deprotonation of D422. A protonated D422 side chain promotes the N−S acyl shift by acting as a hydrogen bond donor to the C1 thiolate, catalyzing step one of protein splicing, while a deprotonated D422 catalyzes transesterification, step two of protein splicing, by activating C+1.[18] *Thus, the deprotonation of the neutral D422 side chain in step one will naturally link step one and step two of protein splicing.* Deprotonation of D422 may be induced by a positive charge nearby, such as the protonated H73 (Figure 5A and B), lowering the D422 pK_a. Thus, the most conserved block-B histidine may also have a coordination role in protein splicing. In addition, pK_a values of C1, H73, and D422 might be coupled and therefore sensitive to the ionization state of each other. Such a network of coupled ionizable groups at the intein active site has likely evolved to ensure the orderly progression of protonation and deprotonation events during catalysis. The deprotonation of D422 may also occur due to the loss of a nearby negative charge such as the thiolate or the oxythiazolidine intermediate, leading to lowered pK_a.

As an aspartate residue occupies the block-F position 4 in more than 60% of canonical inteins, we speculate that the proposed coordination mechanism is likely applicable to more than 200 canonical inteins with a block-F aspartate. We summarize the mechanistic details of step one and two of protein splicing for the *Mtu* RecA intein[14,18,27] in Figure 5.

Step 1, N−S Acyl Shift. The C1 thiol group with a pK_a of 7.5 is activated by the conserved block-B histidine H73 with a pK_a of 7.3 (Figure 5A), as described in our previous paper on H73.[27] The protonated D422 side chain, with a pK_a of 6.1, stabilizes the activated thiolate by serving as a hydrogen bond donor (Figure 5B). The C1 thiolate attacks the scissile bond carbonyl, resulting in the oxythiazolidine intermediate. It is further hypothesized that the protonated H73 or the loss of nearby negative charges will lower D422 pK_a, causing D422 to deprotonate (Figure 5C,D). D422 may give up its proton to a nearby water molecule or to the nascent N-terminal C1 amine, which is formed after H73 protonates the leaving amine group and breaks down the cyclic oxythiazolidine intermediate (Figure 5C). After these reactions, H73 has an extremely low pK_a of <3, due to the nearby positive charged C1 amine,[27] while D422 is in the right ionization state to promote transesterification (Figure 5D).

Step 2, Transesterification and the Formation of the Branched Intermediate. D422 takes a proton from the C+1 thiol group, activating transesterification (Figure 5D), as proposed by Pereira et al.[18] The protonated D422 may stabilize the C+1 thiolate as a hydrogen bond donor, which is consistent with the dramatically depressed C+1 pK_a obtained by Paulus et al.[47]

For step 3, Frutos et al. has demonstrated in the *Mxe* GyrA intein that the formation of the branched intermediate most likely activates the cyclization of C-terminal asparagine and C-terminal cleavage by structural rearrangement of active site residues.[14] Reactions in step 4 are spontaneous and do not require coordination.

In this mechanism, only protonated D422 can initiate protein splicing (Figure 5A). There is evidence that a negatively charged D422 side chain may inhibit C-terminal cleavage, as mutations of D422 to N, A, and G result in substantial C-terminal cleavage.[18] The mechanism of such an inhibition can be electrostatic, due to unfavorable interactions with an oxyanion intermediate during

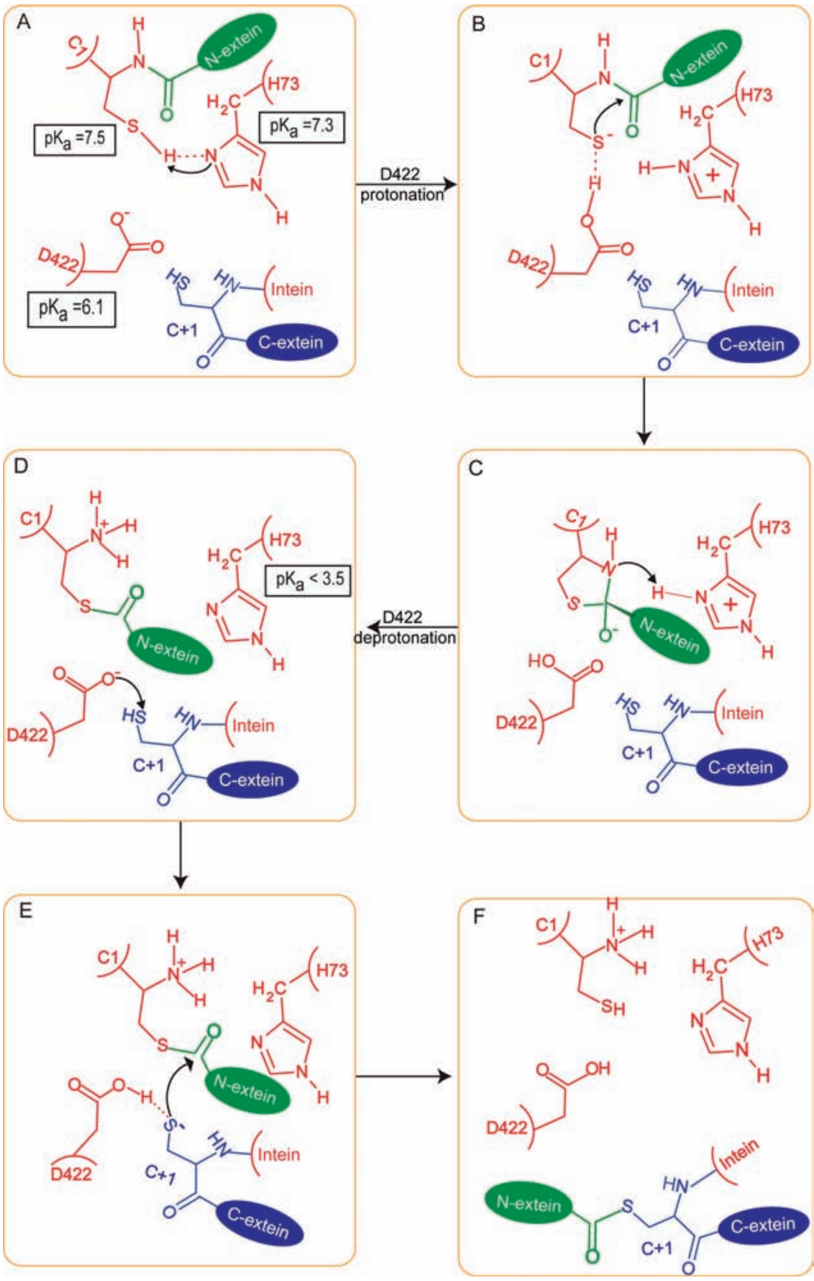


Figure 5. D422 coordination mechanism in *Mtu* RecA for step one and two of protein splicing. Arrows points to direction of electron transfer. (A) Block-B histidine H73 initiates splicing by deprotonating the C1 thiol group. (B) D422 serves as a H-bond bond donor to the C1 thiolate to stabilize the thiolate and facilitates N−S acyl shift. (C) An oxythiazolidine intermediate forms after nucleophilic attack. (D) The nascent amino group lowers the pK_a of H73, while D422 activates the C+1 thiol group. (E) The thiolate group of C+1 carries out nucleophilic attack on the thioester. (F) After the formation of the branched intermediate, succinimide formation and C-terminal cleavage are accelerated, completing the first three steps of protein splicing.

C-terminal cleavage.[48,49] A negatively charged D422 will reduce the efficiency of the N−S acyl shift, as the pK_a of C1 will be higher and C1 thiolate less likely to form. Thus, the protonation of the D422 side chain may serve as a switch to turn on the protein splicing, while the deprotonated D422 side chain minimizes the occurrence of side reactions.

■ CONCLUSION

In this study, we have shown that the block-F D422 has a strikingly high pK_a of 6.1, two units above the normal pK_a of aspartates, indicating a crucial catalytic role for the protonated aspartate side chain. For the first time, we *experimentally* demonstrated the pK_a coupling between two active residues: the elevated pK_a of D422 is coupled to the depressed pK_a of the block-A cysteine C1 in the *Mtu* RecA intein. The pK_a shift coupling and NMR structures demonstrate that the protonated block-F aspartate serves as the hydrogen bond donor to stabilize the C1 thiolate and promote the first step of splicing. Interestingly, in vivo splicing assays with various mutants at the 422 position show that a deprotonated aspartate side chain is also necessary for efficient splicing, most likely due to its deprotonation and activation of the nucleophile at the C-terminal splicing junction and catalyzing step two of protein splicing as shown by Pereira et al.[18] Combining these results, we propose that the sequential protonation and deprotonation of the D422 side chain coordinates the first two steps of protein splicing.

ASSOCIATED CONTENT

Supporting Information. Four figures containing aspartate side chain assignment in ΔΔIhh-SM, pK_a of aspartate in ΔΔIhh-SM, pK_a of aspartate residues in NI, and D422 side chain assignment in CBDΔΔIhh-SM. The atomic coordinates were deposited in the Protein Data Bank, www.pdb.org (PDB code 2L8L). The NMR assignments were deposited in the BioMagResBank (www.bmrb.wisc.edu) under accession number 17414. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
wangc5@rpi.edu.

ACKNOWLEDGMENT

We thank Marlene Belfort for plasmids and reagents. We are grateful to Marlene Belfort and Ken Mills for their critical reading of the manuscript. We acknowledge Nilesh Banavali, Saroj Nayak, Brian Pereira, and Phil Shemella for discussions. We thank support from the NIH (R01GM81408 to C.W.; R01GM44844 to M. Belfort). Y.L. acknowledges the NSF of China (Grant 20873135).

REFERENCES

(1) Paulus, H. *Annu. Rev. Biochem.* **2000**, *69*, 447.
(2) Hirata, R.; Ohsumk, Y.; Nakano, A.; Kawasaki, H.; Suzuki, K.; Anraku, Y. *J. Biol. Chem.* **1990**, *265*, 6726.
(3) Kane, P. M.; Yamashiro, C. T.; Wolczyk, D. F.; Neff, N.; Goebl, M.; Stevens, T. H. *Science* **1990**, *250*, 651.
(4) Tyszkiewicz, A. B.; Muir, T. W. *Nat. Methods* **2008**, *5*, 303.
(5) Mootz, H. D.; Blum, E. S.; Tyszkiewicz, A. B.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 10561.
(6) Blaschke, U. K.; Silberstein, J.; Muir, T. W. *Methods Enzymol.* **2000**, *328*, 478.
(7) Wood, D. W.; Derbyshire, V.; Wu, W.; Chartrain, M.; Belfort, M.; Belfort, G. *Biotechnol. Prog.* **2000**, *16*, 1055.
(8) Paulus, H. *Chem. Soc. Rev.* **1998**, *27*, 375.
(9) Lana Saleh, F. B. P. *Chem. Rec.* **2006**, *6*, 183.
(10) Perler, F. B. *Nucleic Acids. Res.* **2002**, *30*, 383.
(11) Klabunde, T.; Sharma, S.; Telenti, A.; Jacobs, W. R., Jr.; Sacchettini, J. C. *Nat. Struct. Biol.* **1998**, *5*, 31.
(12) Mizutani, R.; Nogami, S.; Kawasaki, M.; Ohya, Y.; Anraku, Y.; Satow, Y. *J. Mol. Biol.* **2002**, *316*, 919.
(13) Mizutani, R.; Anraku, Y.; Satow, Y. *J. Synchrotron Radiat.* **2004**, *11*, 109.
(14) Frutos, S.; Goger, M.; Giovani, B.; Cowburn, D.; Muir, T. W. *Nat. Chem. Biol.* 2010, *6*, 527.
(15) Ding, Y.; Xu, M. Q.; Ghosh, I.; Chen, X.; Ferrandon, S.; Lesage, G.; Rao, Z. *J. Biol. Chem.* **2003**, *278*, 39133.
(16) Wood, D. W.; Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M. *Nat. Biotechnol.* **1999**, *17*, 889.
(17) Van Roey, P.; Pereira, B.; Li, Z.; Hiraga, K.; Belfort, M.; Derbyshire, V. *J. Mol. Biol.* **2007**, *367*, 162.
(18) Pereira, B.; Shemella, P. T.; Amitai, G.; Belfort, G.; Nayak, S. K.; Belfort, M. *J. Mol. Biol.* **2010**, *406*, 430.
(19) Xu, M. Q.; Perler, F. B. *EMBO J.* **1996**, *15*, 5146.
(20) Southworth, M. W.; Amaya, K.; Evans, T. C.; Xu, M. Q.; Perler, F. B. *Biotechniques* **1999**, *27*, 110.
(21) Telenti, A.; Southworth, M.; Alcaide, F.; Daugelat, S.; Jacobs, W. R., Jr.; Perler, F. B. *J. Bacteriol.* **1997**, *179*, 6378.
(22) Amitai, G.; Callahan, B. P.; Stanger, M. J.; Belfort, G.; Belfort, M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11005.
(23) Nogami, S.; Satow, Y.; Ohya, Y.; Anraku, Y. *Genetics* **1997**, *147*, 73.
(24) Mills, K. V.; Paulus, H. *Homing Endonucleases Inteins* **2005**, *16*, 233.
(25) Du, Z.; Liu, Y.; Ban, D.; Lopez, M. M.; Belfort, M.; Wang, C. *J. Mol. Biol.* **2010**, *400*, 755.
(26) Du, Z.; Liu, Y.; Zheng, Y.; McCallum, S.; Dansereau, J.; Derbyshire, V.; Belfort, M.; Belfort, G.; Roey, P.; Wang, C. *Biomol. NMR Assign.* **2008**, *2*, 111.
(27) Du, Z.; Shemella, P. T.; Liu, Y.; McCallum, S. A.; Pereira, B.; Nayak, S. K.; Belfort, G.; Belfort, M.; Wang, C. *J. Am. Chem. Soc.* **2009**, *131*, 11581.
(28) Hiraga, K.; Soga, I.; Dansereau, J. T.; Pereira, B.; Derbyshire, V.; Du, Z.; Wang, C.; Van Roey, P.; Belfort, G.; Belfort, M. *J. Mol. Biol.* **2009**, *393*, 1106.
(29) Hiraga, K.; Derbyshire, V.; Dansereau, J. T.; Van Roey, P.; Belfort, M. *J. Mol. Biol.* **2005**, *354*, 916.
(30) Wang, P.-F.; McLeish, M. J.; Kneen, M. M.; Lee, G.; Kenyon, G. L. *Biochemistry* **2001**, *40*, 11698.
(31) Noda, L. H.; Kuby, S. A.; Lardy, H. A. *J. Am. Chem. Soc.* **1953**, *75*, 913.
(32) Krezel, A.; Bal, W. *J. Inorg. Biochem.* **2004**, *98*, 161.
(33) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289.
(34) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.
(35) Zweckstetter, M. *Nat. Protoc.* **2008**, *3*, 679.
(36) Chou, J. J.; Gaemers, S.; Howder, B.; Louis, J. M.; Bax, A. *J. Biomol. NMR* **2001**, *21*, 377.
(37) Lorieau, J.; Yao, L.; Bax, A. *J. Am. Chem. Soc.* **2008**, *130*, 7536.
(38) Bhattacharya, A.; Tejero, R.; Montelione, G. T. *Proteins* **2007**, *66*, 778.
(39) Won, H.; Kim, J. R.; Ko, K.; Won, Y. *Bull. Korean Chem. Soc.* **2002**, *23*, 27.
(40) Witt, A. C.; Lakshminarasimhan, M.; Remington, B. C.; Hasim, S.; Pozharski, E.; Wilson, M. A. *Biochemistry* **2008**, *47*, 7430.
(41) Pellecchia, M.; Iwai, H.; Szyperski, T.; Wuthrich, K. *J. Magn. Reson.* **1997**, *124*, 274.
(42) Derbyshire, V.; Wood, D. W.; Wu, W.; Dansereau, J. T.; Dalgaard, J. Z.; Belfort, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11466.
(43) Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 388.
(44) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. *Nat. Struct. Biol.* **1996**, *3*, 946.
(45) Assadi-Porter, F. M.; Fillingame, R. H. *Biochemistry* **1995**, *34*, 16186.
(46) Witt, A. C.; Lakshminarasimhan, M.; Remington, B. C.; Hasim, S.; Pozharski, E.; Wilson, M. A. *Biochemistry* **2008**, *47*, 7430.
(47) Shingledecker, K.; Jiang, S.-q.; Paulus, H. *Arch. Biochem. Biophys.* **2000**, *375*, 138.
(48) Sun, P.; Ye, S.; Ferrandon, S.; Evans, T. C.; Xu, M.-Q.; Rao, Z. *J. Mol. Biol.* **2005**, *353*, 1093.
(49) Ding, Y.; Xu, M.-Q.; Ghosh, I.; Chen, X.; Ferrandon, S.; Lesage, G.; Rao, Z. *J. Biol. Chem.* **2003**, *278*, 39133.