Electronic Structure of Neighboring Extein Residue Modulates Intein C-Terminal Cleavage Activity

Philip T. Shemella,[†] Natalya I. Topilina,[¶] Ikko Soga,[¶] Brian Pereira,[‡¶§] Georges Belfort,[‡§] Marlene Belfort,[¶*] and Saroj K. Nayak[†*]

[†]Department of Physics, Applied Physics, and Astronomy, [‡]Howard P. Isermann Department of Chemical and Biological Engineering, and [§]Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York; and [¶]Wadsworth Center, New York State Department of Health, Albany, New York

ABSTRACT   Protein splicing is an autocatalytic reaction where an intervening element (intein) is excised and the remaining two flanking sequences (exteins) are joined. The reaction requires specific conserved residues, and activity may be affected by both the intein and the extein sequence. Predicting how sequence will affect activity is a challenging task. Based on first-principles density functional theory and multiscale quantum mechanics/molecular mechanics, we report C-terminal cleavage reaction rates for five mutations at the first residue of the C-extein (+1), and describe molecular properties that may be used as predictors for future mutations. Independently, we report on experimental characterization of the same set of mutations at the +1 residue resulting in a wide range of C-terminal cleavage activities. With some exceptions, there is general agreement between computational rates and experimental cleavage, giving molecular insight into previous claims that the +1 extein residue affects intein catalysis. These data suggest utilization of attenuating +1 mutants for intein-mediated protein manipulations because they facilitate precursor accumulation in vivo for standard purification schemes. A more detailed analysis of the "+1 effect" will also help to predict sequence-defined effects on insertion points of the intein into proteins of interest.

INTRODUCTION

Protein splicing involves the autocatalytic release of an intervening protein sequence, termed the intein, with the joining of two flanking protein sequences, the exteins (1,2). Experiments have identified key reaction steps in protein splicing and have revealed the conserved residues required for this reaction (1,3–6). It has been shown that in addition to the intein sequence, the extein sequence can also attenuate or accelerate the reaction (7–11).

Understanding these mutations is important because inteins are increasingly used for many biotechnological applications such as purification schemes (12–20), drug development (21,22), molecular sensors (23,24), and eventual industrial biocatalysis (25–28). By mutating the first residue at the N-terminus of the intein from Cys to Ala, the first step of the splicing reaction (N–S acyl shift) is inhibited, thus isolating the C-terminal cleavage step of the reaction (1). (Note that atoms are annotated with one letter—i.e., H = hydrogen—and that amino acids are annotated with three letters—i.e., His = histidine.) For the *Mycobacterium tuberculosis* (*Mtu*) RecA mini-intein, an additional Asp^{422}/Gly mutation creates the C-terminal cleavage mutant (CM), which is extremely active in cleavage, particularly in low pH environments (30–32).

The second step of intein splicing, the transesterification step, usually utilizes Cys, Ser, or Thr as the first residue of the C-extein (+1), with Cys being most prevalent (1). Because the CM was found to be exceedingly reactive at low pH values, Wood et al. (31,32) utilized Met at the first residue of the C-extein, which was the native N-terminus of the protein that formed the C-extein sequence. Note that, in this experiment, three proteins of various sizes were contrasted with only the Cys/Met C-extein mutation: Thymidylate synthase (31.5 kDa), Hfq Protein (18 kDa), and rh aFGF (14 kDa). For these proteins, the Cys-to-Met mutation resulted in a decrease of the reaction rate by a factor of 12.0, 5.0, and 7.8, respectively. They thus observed that Cys/Thr/Ser is not required for C-terminal cleavage but that the residue at +1 is able to modulate the cleavage rate. This result is important for biotechnological applications because for C-terminal cleavage, mutation of the first residue of the C-extein is not constrained.

Given the effect of the +1 residue on C-terminal cleavage, and the impact of this neighboring extein residue on biotechnological applications, we have used first-principles simulations to predict the changes in reactivity due to mutation at the +1 residue. First-principles calculations are based on a quantum mechanical Hamiltonian where each electron is treated explicitly (34). As input, atomic number and total system charge are parameters. From these, the electron wave function can be energetically minimized, and from calculated atomic forces, the geometry may be optimized. From an optimized wave function, we can accurately predict partial atomic charges, molecular orbital

---

energy levels, and the energy difference between states along a reaction pathway.

These properties are of extreme interest for biological systems, where interesting chemistry occurs at transition states instead of at equilibrium (35,36). First-principles calculations, although limited by the system size, are a valuable tool to characterize transition path structures as well as to predict likely reaction mechanisms based on calculated energy barriers. In addition, the same properties such as atomic charge and molecular orbital energies can be calculated at transition structures or at any point along a reaction pathway.

Previously, using similar first-principles calculations, an atomic-level reaction scheme for low pH enhanced C-terminal cleavage was proposed (37). In this mechanism, the amide of the C-terminal scissile peptide bond between Asn and the C-extein +1 residue is protonated by a hydronium ion (H_3O^+), thus making the peptide bond open for nucleophilic attack by the Asn side chain. This reaction then allows for succinimide formation from Asn, and peptide bond cleavage. The active site for C-terminal cleavage is shown in Fig. 1 A. The downstream Cys/Ser/Thr residue at the +1 position of the C-extein has no direct role in the C-terminal cleavage reaction, but does affect the reaction rate (31) by a presently unknown mechanism.

In this study, we use independent computational and experimental mutagenesis studies to investigate the effect on the reaction rate of single amino-acid mutations at the first residue of the C-extein, which flanks the highly conserved His-Asn at the intein C-terminus. To assess activity at the C-terminus of the intein and obtain an atomic-level understanding of the effect of extein mutation on the reaction barrier, detailed quantum mechanical calculations on the intein C-terminal cleavage reaction were carried out for a series of mutants. For our model QM system, mutants with neutral side chains and similar size to the wild-type Cys+1 were considered (Ser, Thr, Ala, Val, and Met; see Fig. 1 B). For our quantum mechanics/molecular mechanics (QM/MM) system, Cys, Ser, Met, and Leu were considered.

The rationale for our computational mutagenesis is as follows: Cys and Met were used in previous experiments (31,32); Ser is similar in size to Cys but more polar; Ala is similar in size to Cys but nonpolar; Val is slightly larger and nonpolar; Leu is similar in size to Met but with hydrophobicity similar to Val; and Thr is similar in size but contains polar and nonpolar elements.

Not considered in this study were: Gly and Pro, which are too small and may distort the tertiary structure; Ile, because of the similarity to Leu and Val; Phe, His, Tyr, and Trp, because they are ring structures which are too large and may have anomalous affects when substituted; Lys, Arg, Glu, and Asp, which have charge and thus are likely to change the local structure too greatly; and Gln and Asn, which may be appropriate to test in future simulations.

Based on our first-principles computational results, the three residues that occur most frequently in nature, Cys,
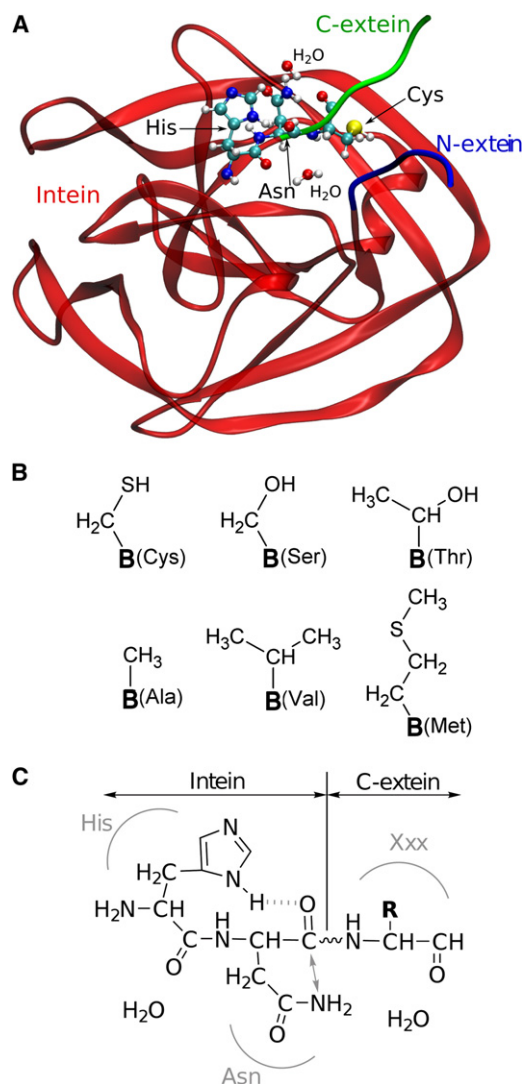

FIGURE 1 The intein and C-extein active site residues. (A) The intein cleavage mutant (CM) crystal structure (PDB code 2IN8) with computationally appended exteins. (B) +1 mutations: side chains for mutants (Cys, Ser, Thr, Ala, Val, and Met). Symbol B represents the protein backbone. (C) Schematic of the His-Asn-Xxx object residues. His and Asn are highly conserved in intein sequences. (Dotted line) H-bond between the His side chain and Asn carbonyl. (Arrow) Cyclization coordinate of Asn. (Wavy line) Scissile peptide bond between C(Asn) and N(Xxx). R represents the +1 side chain.

Ser, and Thr, are the fastest, while Val, Ala, Leu, and Met attenuate the reaction rate.

A novel cleavage assay (38) was used to assess activity at the C-terminus of the intein, with Cys at +1 replaced by the same five amino acids (Fig. 1 B). In approximate accord with the computational assays, modulation was observed relative to the wild-type Cys, and naturally occurring amino acids Ser and Thr were more active than Ala, Met, and Val. The results obtained herein suggest the possibility of controlling the intein cleavage reaction by mutation of the +1 residue. Specifically, we suggest the utility of the

Cys+1/Val mutant, sometimes abbreviated to Val+1, for intein-based purification schemes. In addition to helping control the kinetics of cleavage, modeling the behavior of the +1 extein residue will help to predict constraints on where an intein is inserted into a protein of interest, thus increasing the utility of inteins as components of biotechnological applications.

METHODS

Computational methods

Computational methods include first-principles density functional theory (39,40), used to calculate structure and energy profiles for stable, intermediate, and transition states. In particular, we have used the widely implemented Becke hybrid functional (41), B3LYP, with Gaussian code (42). Large basis sets [6-31G(d,p) and 6-311++g(d,p)] were used for first locating and then precisely calculating the energy barriers, respectively.

In addition to QM calculations, the multiscale quantum mechanics/molecular mechanics (QM/MM) method was employed. The protein active site and critical solvent molecules were treated with first-principles methods, whereas the remaining full-protein system was considered with classical potentials (35,36,43–47). Based on the intein crystal structure for the (*Mtu*) RecA intein ($\Delta\Delta I_{hh}$-CM PDB code No. 2IN8) (48), abbreviated N- and C-terminal exteins were computationally added (49). The N-extein sequence consisted of Ace-Val-Val-Lys-Asn-Lys, where Ace is an acetyl-beginning group for the N-extein. The C-extein sequence consisted of Cys-Ser-Pro-Pro-Phe-Nme, where Nme is an N-methylamine ending group for the C-extein. Both exteins are based on the native extein sequences. Classical equilibration of protein and solvent were carried out for 4 ns with temperature of 298 K, pressure of 1 bar, and 9548 and 9549 water molecules for the Cys and Met systems, respectively. The full protein and solvent system was equilibrated with classical molecular dynamics simulations for 4 ns with the standard AMBER force field (50). Then, the complete classical system was trimmed down to include the protein (intein and exteins) as well as all interior waters and those exterior water molecules within a range of 7.0 Å to the protein surface. For each QM/MM calculation, all atoms were relaxed, and each calculation included at least 6500 atoms.

Experimental methods: intein fusion and cleavage assay

Escherichia coli BL21(DE3) star (Invitrogen, Carlsbad, CA) containing the intein cleavage construct (HT:intein:GFP (where HT is His-tag; GFP is green fluorescent protein)), cloned into pET30b, was used to overexpress the intein fusion (38) (Fig. 2 *A*). In the HT:intein:GFP construct, the (*Mtu*) RecA $\Delta\Delta I_{hh}$ mini-intein (51) is flanked by a synthetic HT N-extein and a C-extein comprising super-folder GFP (Fig. 2 *A*). The (*Mtu*) RecA mini-intein contained mutations: Cys1/Ala to inhibit N-terminal cleavage and splicing (52), Val[67]/Leu (38,53), and Asp[422]/Gly to create the cleavage mutant (CM) (30).

Cells containing HT:intein:GFP were grown overnight at 37°C, diluted 100-fold, grown at 37°C to OD$_{600}$ ~0.5, and protein expression was induced as previously described (38). Cultures were incubated with shaking at 25°C for 1 h, and cells were harvested by centrifugation. The cells were then resuspended in cleavage reaction buffer (0.5 mM NaCl, 1 mM EDTA, and 50 mM sodium phosphate, pH 6.0) in volume equal to 1:5 volume of cell culture and sonicated on ice. Lysates were incubated at 19°C, 25°C, 30°C, and 37°C for the indicated times, and 10 μg protein was loaded onto a 12% SDS-PAGE gel, without boiling (Fig. 2 *B*). All experiments were done in triplicate.

To test the accumulation of the Cys+1/Val precursor in vivo, cells containing HT:intein:GFP with wild-type Cys+1 or mutant Val+1 were grown
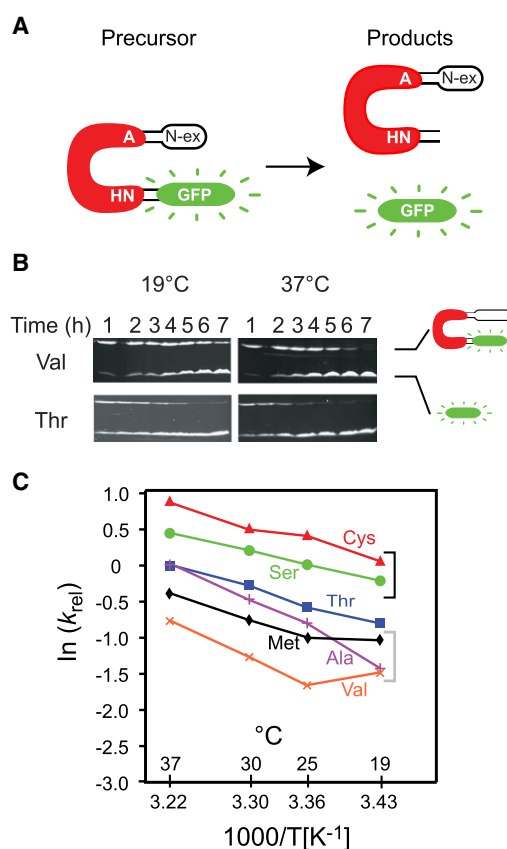

FIGURE 2 Experimental cleavage assay. (*A*) Precursor (HT:intein:GFP) and GFP cleavage product are shown. (*B*) Gel analysis of precursor and cleavage product. A time-course is presented, indicating the identity of the two bands at two different temperatures for the Cys+1/Val and Cys+1/Thr mutants. (*C*) Experimental cleavage rates are shown for +1 residues Cys, Ser, Thr, Ala, Val, Met and various temperatures, relative to Cys, at 19°C.

overnight at 37°C, diluted 100-fold, grown at 37°C to OD$_{600}$ ~0.5, and protein expression was induced by 0.4 mM isopropyl-β-D-thio-galactoside (38) at 20°C for 1, 2, 3, 4, and 20 h. Cells were harvested by centrifugation, and then resuspended in cleavage reaction buffer in volume equal to 1:5 (1 and 2 h inductions), 1:2.5 (3 and 4 h inductions), and 1:1.25 (20 h induction) volume of cell culture and sonicated on ice.

Fluorescence signal was detected with 365-nm ultraviolet light on a FluroChem-8900 (Alpha Innotech, San Leandro, CA), revealing two bands—the HT:intein:GFP precursor and the GFP product (Fig. 2 *B*). Quantitation of the bands was by spot densitometry using AlphaEase FC (Alpha Innotech) software.

To test the stability of the Cys+1/Val precursor in vitro, cell lysates or the Ni-NTA-purified precursor (with cleavage products) were incubated at 4°C for 1, 2, 3, 4, and 20 h. The proteins were separated by polyacrylamide gel electrophoresis and the precursor and cleavage product were visualized by GFP fluorescence.

RESULTS

Computational studies provide atomic rationale for differences in k_{rel} of mutants

To gain an atomic-level understanding of the kinetic modulation of C-terminal cleavage activity, quantum mechanical
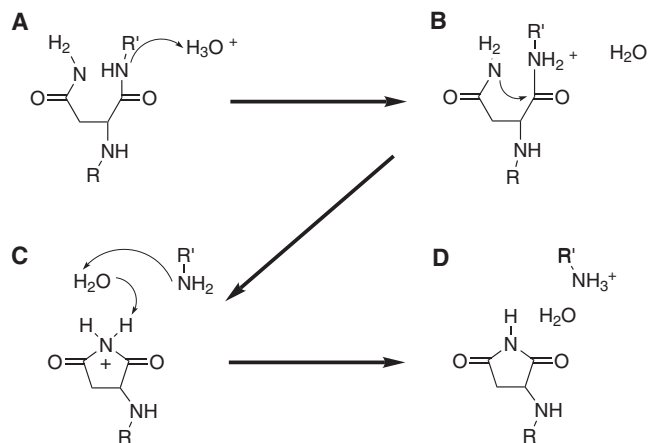
FIGURE 3   Proposed N-protonation reaction scheme for Asn cyclization to succinimide. R and R' represent the intein and C-extein, respectively. (*A*) Precursor and hydronium ion. (*B*) N-protonation step. (*C*) Asn cyclization. (*D*) Succinimide formation and C-terminal cleavage.

calculations were performed with five individual mutations at the Cys+1 extein residue: Thr, Ser, Ala, Val, or Met. Simulations were based on both full quantum mechanical molecular analysis, and a hybrid quantum mechanics and molecular mechanics (QM/MM) multiscale approach.

As a model system, a tripeptide active site system (His-Asn-Cys) was used in our computational study, and is shown in Fig. 1 *C*. The tripeptide system does not include external interactions but is a useful starting point to be later validated with QM/MM. The energy barriers for isolated systems such as this tripeptide system are often exaggerated due to the lack of a dielectric background, but these calculations with reduced external interactions are critical to understand the underlying phenomena of the chemical reaction. Intein crystal structures usually include a hydrogen bond between the H^{δ} of the penultimate His side chain and the carbonyl O of Asn (shown in Fig. 1 *C* as a *dotted line*), the final amino acid of the intein (3,4,54–56). Although the penultimate His residue was previously assumed the proton donor for the C-terminal cleavage reaction in the context of splicing (5), further inspection revealed that this was not the case for pH-dependent C-terminal cleavage. For a simple proton-catalyzed reaction, there is an inverse linear rate dependence on the pH, which was observed experimentally for the C-terminal cleavage reaction (31). Because the ability of His to act as an acid is based on its local pK_a value, the expected pH rate-curve should be nonlinear, specifically sigmoidal in shape, which is in contrast to the linearity observed experimentally (30,31).

A mechanism for C-terminal cleavage was proposed that explained the increased reaction rate at low pH (37,57,58), and has been used for this study (Fig. 3 *A*). Recently, strain in the peptide bond dihedrals was shown to increase the propensity for protonation of the peptide amide (59). After N-protonation occurred via a hydronium ion (H_3O^+), the

peptide bond was elongated, resonance between C-N was lost, the nitrogen was sp^3 hybridized, and the carbonyl carbon became open for attack by the Asn side chain (60) (Fig. 3 *B*). The Asn side chain underwent cyclization into succinimide (Fig. 3 *C*) and the peptide bond was irreversibly cleaved (Fig. 3 *D*).

The His-Asn-Cys tripeptide system (Fig. 1 *C*) was studied with various mutants in place of the wild-type Cys, shown in Fig. 1 *B*. From previous experimental results, Met was known to attenuate the reaction rate (31,32). We have considered Ser, Thr, Ala, Val, and Met with high-level QM methods and then compared the computational reaction rates with those determined experimentally.

For all mutants, the hydrogen bond between H^{δ} of His and the carbonyl O of Asn (*dashed line* in Fig. 1 *C*) caused O to be energetically unable to accept a proton from H_3O^+. This hydrogen bond is usually found at the C-terminus of inteins and is important for reducing the possibility of proton transfer to the carbonyl O. In fact, the normally exothermic reaction for H_3O^+ to donate a proton to the carbonyl O atom is endothermic for cases where O was hydrogen-bonded to another group (57).

Computational energy barriers and relative rate constants are shown in Table 1 for the chemical reaction that occurs between states B and C in Fig. 3. For the wild-type His-Asn-Cys, the computational energy barrier in the gas phase was 27.95 kcal/mol, using the N-protonation mechanism calculated with the tripeptide system, in reasonable agreement with the experimental results of ~21 kcal/mol (31). Ser and Thr mutants had similar energy barriers, suggesting similar activity to the wild-type for this group of mutants (1.40 times as active for Ser, 1.93 for Thr). Ala, Val, and Met had increased barriers, indicating a slower reaction rate for this group (0.31 times as active for Ala, 0.17 for Val, and 0.06 for Met).

TABLE 1   Computational energy barriers (ΔE_i) for various C-extein mutations (His-Asn-Xxx), and computational relative reaction rates (k_{rel} and $ln(k_{rel})$)

	Mutant (Xxx)	(ΔE_i) [kcal/mol]	k_{rel}	$ln(k_{rel})$
QM	Cys	27.95	1	0
	Thr	27.56	1.93	0.65
	Ser	27.75	1.40	0.33
	Ala	28.64	0.31	−1.16
	Val	28.97	0.17	−1.72
	Met	29.58	0.06	−2.75
QM/MM	Cys	26.17	1	0
	Ser	25.53	2.94	1.08
	Met	27.07	0.21	−1.51
	Leu	28.28	0.02	−3.56

Reaction rates for quantum mechanics (QM) are relative to the His-Asn-Cys wild-type at $T = 25°C$, and for quantum mechanics/molecular mechanics (QM/MM) to the full protein system, and are plotted with experimental results in Fig. 2 *C*. The Arrhenius equation was used to compare mutants: $k_{rel} = k_1/k_2 = e^{-(\Delta E_1 - \Delta E_2)/RT}$, where k_i and ΔE_i are the reaction rate and energy barrier for the ith mutant, respectively, and R is the gas constant and T is the temperature in Kelvin.

Cleavage assays indicate a hierarchy of rate constants

We made the same five individual mutations at the Cys+1 extein position in the fused intein cleavage construct (HT:intein:GFP), containing a short N-extein, an intein, and a C-extein comprising green fluorescent protein (Fig. 2 A). The intein contains a Cys1/Ala mutation to inhibit N-terminal cleavage and splicing, and an Asp^{422}/Gly mutation to enhance C-terminal cleavage. The "effect of" C-extein Cys+1 residue "substitution" was assayed for cleavage over 24 h on gels (Fig. 2 B), and the percent cleavage were quantified. The data were modeled as a first-order exponential for the appearance of cleavage product normalized with respect to total GFP signal, according to

$$G_t = G_0 + P_0\left(1 - e^{-kt}\right), \qquad (1)$$

where G_t = GFP product at time t, G_0 = GFP at t_0, and P_0 = precursor (active HT:intein:GFP) at t_0. The data were analyzed over triplicate experiments and the model was fit to each set of data (38) (see Fig. S1 in the Supporting Material). Rate constants (k) were derived from the fit and are shown in Fig. 2 C. The experimental results indicated that a wide range of rate constants for C-terminal cleavage can occur in mutants of Cys+1. The native Cys gave the highest rate constant for all the temperatures tested, between 19°C and 37°C. The other two residues found in nature, Ser and Thr, yielded rate constants that were 50–10% that of Cys. The three nonnative residues, Ala, Met, and Val, had lower rate constants with Val having a rate <5% of Cys.

Additional cleavage assays were considered that focused on C-extein mutations at C+1 with hydrophobic side chains. Reaction rates were determined at $T = 25°C$, and similar to other experimental rates shown in Fig. 2 C and computational rates in Table 1, when Cys+1 was mutated to a hydrophobic side chain there was strongly attenuated cleavage. In particular, Cys+1/Val the rate decreased by more than an order of magnitude ($k_{rel} = 0.04$). For both Leu and Ile mutants, the reaction was measured to be ~0.07 as fast as the wild-type Cys.

Experimentally and computationally determined rate constants are in general agreement

With a few exceptions, the experimental results are in good agreement with the computationally determined rate constants. In particular, there is correspondence in the grouping of faster mutants (Cys, Ser) and slower mutants (Met, Ala, Val). These results indicate that the side chain of the first C-extein residue is indeed important for C-terminal cleavage kinetics. Interestingly, the two hydrophobic residues (Val and Ala) have breaks in the Arrhenius plots between 25 and 19°C (Fig. 2 C), which may be due to cold-sensitive hydrophobic interactions.

The first discrepancy between theory and experiment is the relative reaction rate of Met at the +1 position. In the experimental findings of this work, Met is still an attenuating mutant although the reaction is considerably more active than the computational prediction. Interestingly, Wood et al. (31,32) found that the rate is decreased by approximately an order of magnitude ($ln(0.1) = -2.30$), in excellent agreement with simulation results. A possible explanation for the discrepancy between the current and previous experimental results is the use of different exteins beyond just the Cys+1/Met mutation. We have indeed demonstrated recently that extein residues can exert dramatic effects on intein function (11).

The second difference involves the Thr mutant, which is predicted by simulation to be the most active, whereas experimental cleavage rates indicate intermediate activity. (Note that Cys, Ser, or Thr is required for the transesterification step of splicing, and of the 486 total inteins in the NEB InBase (the Intein Database; http://www.neb.com/neb/inteins.html), 192 contain Cys+1 (39.5%), 168 have Ser+1 (34.6%), and 107 have Thr+1 (22.0%), whereas other residues account for only 19 inteins (0.04%).) The nucleophilic and polar side chains of Ser and Thr are able to form hydrogen bonds with the backbone groups of a protein, and this results in weaker solvation energies of the side chain than even the nonpolar Gly (62). Unlike Ser, Thr contains an additional nonpolar methyl group, and this may disrupt the solvent properties near the Thr side chain and thus may attenuate the reaction by acting more like Ala than Ser. We therefore propose that the Thr side chain at the C-extein +1 position can act in dual roles: either as a polar –OH group, similar to Ser, or as a nonpolar $-CH_3$ group, similar to Ala. Without a crystal structure that includes the Thr+1 mutant, the local solvent structure cannot be easily predicted. Indeed, our tripeptide model system allows for free rotation of the side chain, something not possible in the context of a folded protein, and a promising subject for a future study.

Multiscale methods (QM/MM) confirm energy barrier differentiation between mutants

The entire protein (Fig. 1 A) and solvent are treated classically with parameterized force fields in a molecular mechanics (MM) calculation, while the active site is treated with QM methods. The full protein QM/MM reaction profile was also calculated (2346+ protein atoms, 4161 water atoms) with the QM active site region containing the same tripeptide region His-Asn-Xxx as well as two water molecules (53+ total QM atoms) (63), to better understand the properties of the +1 mutants. For the wild-type Cys+1, Fig. 4 shows the full QM/MM reaction energy profile with and without electrostatic embedding. The energy barrier was 24.96 kcal/mol for the QM/MM calculation with geometry optimization, in excellent agreement with the 21 kcal/mol measured experimentally (31).
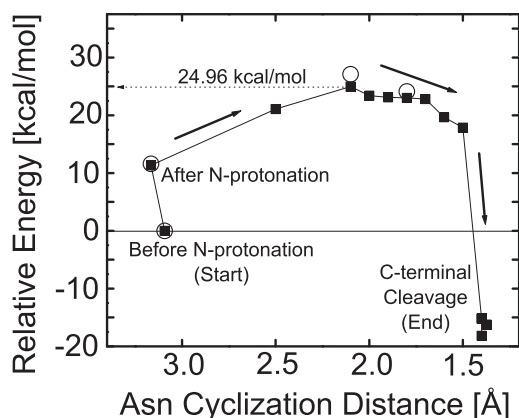
FIGURE 4 QM/MM reaction energy profile for His-Asn-Cys plus two-water QM system. QM/MM geometry optimization (■). QM/MM + charge embedding single point energies (○).

Using the B3LYP/6-31G(d,p) level of theory, independent and parallel reaction profiles for the Met, Cys, Ser, and Leu mutations were calculated using a different QM/MM starting structure. Because Met and Leu are larger in size than Cys, we ran additional classical molecular dynamics equilibration with Met+1. Then, it was straightforward to mutate Met+1 to Cys, Ser, or Leu, which are either smaller or of similar size. Because the starting point is slightly different, and because full protein equilibration was done with a larger and nonpolar side chain, there are small changes to the energy barrier for Cys compared to the initial QM/MM case. For Met, the barrier was 27.07 kcal/mol, for Cys was 26.17 kcal/mol, for Ser was 25.53 kcal/mol, and for Leu was 28.28 kcal/mol (QM/MM energy barriers are included in Table 1). With Met (as part of the QM/MM system), the energy barrier was 0.90 kcal/mol higher than with the wild-type Cys. To compare computational mutation activities with experimental ones, we consider the difference between computational energy barriers and its logarithm—values that can be directly related to relative reaction rates of experiments. (Note that because a mutation at +1 may cause the reaction rate to be greatly modified, and to enable comparison between simulation and experiment, the log of the reaction rate is used.) This difference between energy barriers corresponded to relative reaction rates of

$$ln\big(k_{Cys}/k_{Met}\big) = ln(0.21) = -1.51.$$

This is in good agreement with experimental results of Wood et al. (31), consistent with the results from the smaller computational study (without QM/MM), and in better agreement with the present experimental relative reaction rate for the Met mutant (Fig. 2 C). The slight decrease in energy barrier when Cys was mutated to Ser was consistent with the results of the tripeptide model system. From our computational results, and in agreement with our experimental results, Ser was not an attenuating mutant. This is

of interest due to the necessity of Cys or Ser not only in the C-terminal cleavage reaction, but in the overall intein splicing reaction (1,2). For Leu, the predicted reaction rate was decreased as compared to Cys and was in good agreement with experimental findings for Leu and other hydrophobic side chains (see Table 1).

By increasing the size of our computational system, we are better able to model and understand the effect of the +1 mutant on the reaction rate and to make comparisons to experiment. Because our computational exteins are abbreviated, and do not include either the mechanical constraints that may exist for full proteins or the possibility of those extein-proteins to dock, our computational predictions are expected to be somewhat qualitative. Nevertheless, we are able to group mutants as "faster" or "slower", and this is in agreement with experimental results. The discrepancy that persists between QM/MM and experiment suggests future calculations that may include dynamic effects due to structural differences between intein product and precursor proteins, as well as improved accuracy in predicting very small energy differences between reaction steps for mutants. We expect that our work will stimulate further computational and experimental work in this direction, and we are confident that with future computational power and improved methodology, simulations will play an ever-increasing role in predictions and thereby be an invaluable guide to experimental approaches.

Contrasting the electronic structure of mutants shows distinct properties

To understand the effect of the mutation of the first C-extein amino-acid side chain on the energy barrier, the isolated Cys and Met amino-acid molecules were studied. The electron affinities (EA) and ionization potentials for each were calculated with the B3LYP/6-311++G(d,p) level of theory. The EA for Cys (the amount of energy gained or lost when the system went from neutral to negatively charged) was +6.79 kcal/mol. For Met, the EA was +8.27 kcal/mol, signifying that the side chain of the gas phase Cys residue was more readily able to accept an electron than Met. The reason that Cys was more stable with charge than Met was due to the unique electronic structure and bonding for each S atom. Although each side chain contained an S atom, for Cys the S atom was bonded to one methyl group and one H atom. For Met, both bonds of the S atom were to C atoms. Unique electron occupation and partial atomic charges were calculated with natural population analysis (65).

In changing from neutral to negatively charged, the partial charge of S for Cys went from –0.01 to –0.12 units of charge, corresponding to the addition of 0.11 electrons. For Met, the charge went from +0.17 to +0.13 units of charge, corresponding to the gain of only 0.04 electrons. The S of Cys was able to accommodate more than twice the amount of delocalized electron population as compared to Met, indicating more stability that is energetic in the
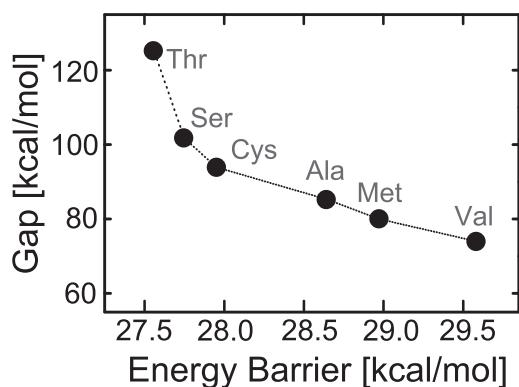
FIGURE 5 Energy difference between the highest occupied molecular orbital (HOMO) for the neutral system and negatively charged system, shown for the isolated amino-acid molecules (Thr, Ser, Cys, Ala, Val, Met) and plotted by their computational energy barrier in the tripeptide system.

negatively charged system. The difference in ionization potential for the same isolated Cys and Met amino acids was calculated. The removal of one electron from Cys required +203.05 kcal/mol whereas that for Met was +191.14 kcal/mol. Because Met was more stable when an electron was removed, and Cys was more stable when an electron was added, we conclude that the electron-pulling and electron-pushing properties, or polarizability, of the first C-extein amino-acid side chain may have an effect on the properties of the scissile peptide bond.

For the isolated amino acids (Thr, Ser, Cys, Ala, Val, and Met), the highest occupied molecular orbitals (HOMO) for the neutrally charged system as well as the negatively charged system were compared. The difference in energy between the HOMO of the negatively charged and the neutral system is termed the energy gap (Fig. 5), and is a measure of how readily the amino acid accepts an electron. As the energy barrier increased for a particular mutant (Ala, Val, Met), the energy gap decreased, implying less favorable conditions for accepting an electron. Those mutants were independently found to attenuate C-terminal cleavage

activity experimentally (Fig. 2 C). The correspondence between the energy gap and the reaction energy barrier is of interest because the electronic structure of an isolated molecule representing an amino-acid side chain may be used to explain and perhaps predict the relative reaction rate for an unknown mutant at the first C-extein position.

Biotechnological applications of Cys+1 mutants

The attenuating mutations, and in particular Val+1, may be useful for future intein-mediated purification schemes. With Cys+1, the C-terminal cleavage reaction was 70% completed in vivo when cells were induced at 25°C, but with Val+1, the reaction was only 30% completed (Fig. S1). Because of the potential utility of the Val+1 mutant in biotechnological applications and the pronounced temperature dependence observed for cleavage reaction (Fig. 2 C), we wished to compare the Val+1 precursor build-up with that of the Cys+1 wild-type when protein synthesis was induced at lower temperature. The in vivo accumulation of the Cys+1/Val precursor was studied by induction of Cys+1 and Val+1 precursor synthesis at 20°C for various times and analysis of the progression of in vivo cleavage. Unlike the wild-type parent, the Cys+1/Val mutant precursor accumulated to high levels with time (Fig. 6 A) and, importantly, >90% of the intact precursor could be recovered under all conditions tested (Fig. 6 A).

The recovered Cys+1/Val mutant precursor is stable in vitro as confirmed by incubation of the precursor either in cell lysate or as a Ni-NTA-purified protein at 4°C over 24 h (Fig. S2). To determine whether time of induction affected cleavage, precursors from 1 h, 4 h, and 20 h postinduction were incubated in vitro. Cleavage of the Cys+1/Val precursor proceeded to completion within 20 h of incubation at 37°C in cleavage buffer (Fig. 6 B), with >80% of the cleavage completed within the first 4–5 h of incubation (Fig. 2 B). The stability of the Val+1 mutant suggests its viability for protein purification purposes. Indeed, when
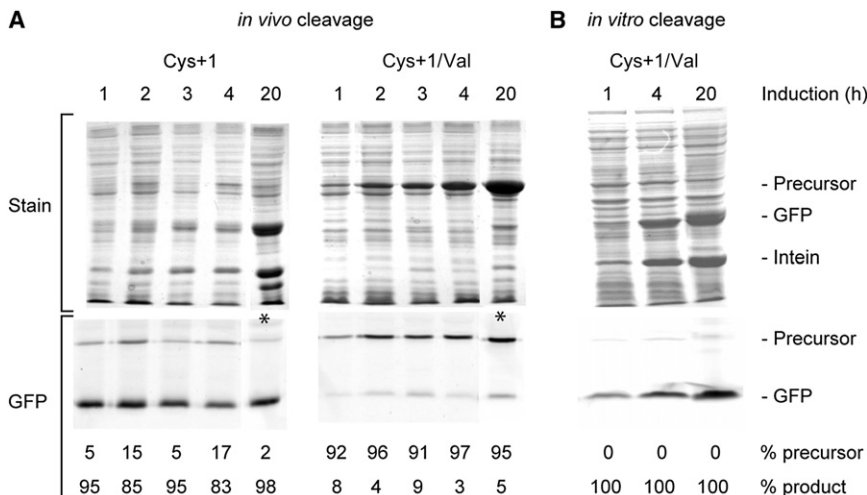

FIGURE 6 Precursor accumulation and cleavage in the Val+1 mutant. (A) Accumulation of precursor of the Cys+1/Val mutant in vivo. Gel analysis of the accumulation of the Cys+1 and Cys+1/Val precursor and cleavage products from cells induced for various times is shown. Proteins were visualized by Coomassie stain (top), and by GFP fluorescence (bottom). The intensity of bands in the lanes with protein induction for 20 h was adjusted to illustrate the ratio of band intensities (asterisks). (B) Cleavage products of the Cys+1/Val mutant in vitro. Lysates from 1 h, 4 h, and 20 h inductions were incubated at 37°C for 20 h. Precursor and cleavage products visualized by Coomassie stain and by GFP fluorescence as in symbol A.

making a Cys+1/Val mutation in a construct previously designed for protein purification (66), the desired product, human acidic fibroblast growth factor (aFGF), was purified to homogeneity in a single step (Fig. S3).

Rapid cleavage of the wild-type intein in vivo hampers precursor isolation for both research and biotechnological applications. For example, obtaining crystal structures of a wild-type intein with exteins attached is one of the outstanding goals of the field. Thus, by modulating splicing rates, both research and application will be advanced through insights provided by our study.

CONCLUSIONS

Inteins are important for biotechnological applications because they catalyze rearrangements of peptide bonds to their adjacent extein sequences. These extein sequences in turn influence reaction kinetics that is important for such applications as intein-mediated peptide ligation and cleavage for protein purification. Using in silico site mutagenesis, we have shown a hierarchy of reaction rates for C-terminal cleavage where the wild-type Cys+1 was replaced with Ser, Thr, Ala, Val, or Met. Our independent experimental mutagenesis study of C-terminal cleavage activities for the same set of mutations agreed well with computational trends with some quantitative differences. Experimental assays show that the +1 mutation modulated the C-terminal cleavage rate, yet the cleavage reaction went to completion in all cases.

The data presented suggest that +1 mutants, especially Cys+1/Val, may be useful for intein-based purification schemes. The general agreement between computational and experimental results can lead to future implementation of theoretical treatment of intein manipulations for development of intein-based devices with designed, adaptable, and tunable kinetic properties.

SUPPORTING MATERIAL

REFERENCES

1. Paulus, H. 2000. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* 69:447–496.

2. Perler, F. B. 2000. InBase, the intein database. *Nucleic Acids Res.* 28:344–345.

3. Klabunde, T., S. Sharma, …, J. C. Sacchettini. 1998. Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat. Struct. Biol.* 5:31–36.

4. Duan, X. Q., F. S. Gimble, and F. A. Quiocho. 1997. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell.* 89:555–564.

5. Ding, Y., M. Q. Xu, …, Z. Rao. 2003. Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J. Biol. Chem.* 278:39133–39142.

6. Clarke, N. D. 1994. A proposed mechanism for the self-splicing of proteins. *Proc. Natl. Acad. Sci. USA.* 91:11084–11088.

7. Chong, S., G. E. Montello, …, J. Benner. 1998. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res.* 26:5109–5115.

8. Chong, S., K. S. Williams, …, M. Q. Xu. 1998. Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. *J. Biol. Chem.* 273:10567–10577.

9. Southworth, M. W., K. Amaya, …, F. B. Perler. 1999. Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *Biotechniques.* 27:110–114, 116, 118–120.

10. Iwai, H., S. Züger, …, P. H. Tam. 2006. Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.* 580:1853–1858.

11. Amitai, G., B. P. Callahan, …, M. Belfort. 2009. Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. USA.* 106:11005–11010.

12. Amitai, G., and S. Pietrokovski. 1999. Fine-tuning an engineered intein. *Nat. Biotechnol.* 17:854–855.

13. Chong, S., F. B. Mersha, …, M. Q. Xu. 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene.* 192:271–281.

14. Wu, W., D. W. Wood, …, M. Belfort. 2002. Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res.* 30:4864–4871.

15. Miao, J., W. Wu, …, G. Belfort. 2005. Single-step affinity purification of toxic and non-toxic proteins on a fluidics platform. *Lab Chip.* 5:248–253.

16. Banki, M. R., L. A. Feng, and D. W. Wood. 2005. Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nat. Methods.* 2:659–661.

17. Banki, M. R., T. U. Gerngross, and D. W. Wood. 2005. Novel and economical purification of recombinant proteins: intein-mediated protein purification using in vivo polyhydroxybutyrate (PHB) matrix association. *Protein Sci.* 14:1387–1395.

18. Sharma, S. S., S. Chong, and S. W. Harcum. 2006. Intein-mediated protein purification of fusion proteins expressed under high-cell density conditions in *E. coli. J. Biotechnol.* 125:48–56.

19. Wang, L., J. H. Kang, …, E. K. Lee. 2010. Expression of intein-tagged fusion protein and its applications in downstream processing. *J. Chem. Technol. Biotechnol.* 85:11–18.

20. Fong, B. A., W. Y. Wu, and D. W. Wood. 2010. The potential role of self-cleaving purification tags in commercial-scale processes. *Trends Biotechnol.* 28:272–279.

21. Paulus, H. 2003. Inteins as targets for potential antimycobacterial drugs. *Front. Biosci.* 8:s1157–s1165.

22. Cheriyan, M., and F. B. Perler. 2009. Protein splicing: a versatile tool for drug discovery. *Adv. Drug Deliv. Rev.* 61:899–907.

23. Mootz, H. D., and T. W. Muir. 2002. Protein splicing triggered by a small molecule. *J. Am. Chem. Soc.* 124:9044–9045.

24. Muralidharan, V., and T. W. Muir. 2006. Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Methods.* 3:429–438.

25. Schmid, A., J. S. Dordick, …, B. Witholt. 2001. Industrial biocatalysis today and tomorrow. *Nature.* 409:258–268.

26. Lesaicherre, M. L., R. Y. P. Lue, …, S. Q. Yao. 2002. Intein-mediated biotinylation of proteins and its application in a protein microarray. *J. Am. Chem. Soc.* 124:8768–8769.

27. Egorova, K., and G. Antranikian. 2005. Industrial relevance of thermophilic Archaea. *Curr. Opin. Microbiol.* 8:649–655.

28. Myung, S., Y. Wang, and Y. H. P. Zhang. 2010. Fructose 1,6-bisphosphatase from a hyper-thermophilic bacterium *Thermotoga maritima*: characterization, metabolite stability, and its implications. *Process Biochem.* 45:1882–1887.

29. Reference deleted in proof.

30. Wood, D. W., W. Wu, …, M. Belfort. 1999. A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotechnol.* 17:889–892.

31. Wood, D. W., V. Derbyshire, …, G. Belfort. 2000. Optimized single-step affinity purification with a self-cleaving intein applied to human acidic fibroblast growth factor. *Biotechnol. Prog.* 16:1055–1063.

32. Wood, D. W. 2000. Generation and application of a self-cleaving protein linker for use in single-step affinity fusion based protein purification. PhD thesis, Rensselaer Polytechnic Institute, Troy, NY.

33. Reference deleted in proof.

34. Cramer, C. J. 2004. Essentials of Computational Chemistry: Theories and Models. John Wiley & Sons, New York.

35. Senn, H. M., and W. Thiel. 2009. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. Engl.* 48:1198–1229.

36. Ranaghan, K. E., and A. J. Mulholland. 2010. Investigations of enzyme-catalyzed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods. *Int. Rev. Phys. Chem.* 29:65–133.

37. Shemella, P., B. Pereira, …, S. K. Nayak. 2007. Mechanism for intein C-terminal cleavage: a proposal from quantum mechanical calculations. *Biophys. J.* 92:847–853.

38. Hiraga, K., I. Soga, …, M. Belfort. 2009. Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. *J. Mol. Biol.* 393:1106–1117.

39. Hohenberg, P., and W. Kohn. 1964. Inhomogeneous electron gas. *Phys. Rev. B.* 136:864–871.

40. Kohn, W., and L. J. Sham. 1965. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140:1133–1138.

41. Becke, A. D. 1993. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98:5648–5652.

42. Frisch, M. J., G. W. Trucks, …, J. A. Pople. 2004. Gaussian 03, Rev. C.02. Gaussian, Wallingford, CT.

43. Maseras, F., and K. Morokuma. 1995. IMOMM: a new integrated ab initio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* 16:1170–1179.

44. Gao, J., and D. G. Truhlar. 2002. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* 53:467–505.

45. Cui, Q., M. Elstner, and M. Karplus. 2002. A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J. Phys. Chem. B.* 106:2721–2740.

46. Torrent, M., T. Vreven, …, H. B. Schlegel. 2002. Effects of the protein environment on the structure and energetics of active sites of metalloenzymes. ONIOM study of methane monooxygenase and ribonucleotide reductase. *J. Am. Chem. Soc.* 124:192–193.

47. Vreven, T., K. Morokuma, …, M. J. Frisch. 2003. Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.* 24:760–769.

48. Van Roey, P., B. Pereira, …, V. Derbyshire. 2007. Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *J. Mol. Biol.* 367:162–173.

49. Davis, E. O., S. G. Sedgwick, and M. J. Colston. 1991. Novel structure of the recA locus of *Mycobacterium tuberculosis* implies processing of the gene product. *J. Bacteriol.* 173:5653–5662.

50. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

51. Hiraga, K., V. Derbyshire, …, M. Belfort. 2005. Minimization and stabilization of the *Mycobacterium tuberculosis* recA intein. *J. Mol. Biol.* 354:916–926.

52. Du, Z., P. T. Shemella, …, C. Wang. 2009. Highly conserved histidine plays a dual catalytic role in protein splicing: a pK_a shift mechanism. *J. Am. Chem. Soc.* 131:11581–11589.

53. Du, Z., Y. Liu, …, C. Wang. 2010. Backbone dynamics and global effects of an activating mutation in minimized Mtu RecA inteins. *J. Mol. Biol.* 400:755–767.

54. Mizutani, R., S. Nogami, …, Y. Satow. 2002. Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides. *J. Mol. Biol.* 316:919–929.

55. Poland, B. W., M. Q. Xu, and F. A. Quiocho. 2000. Structural insights into the protein splicing mechanism of PI-SceI. *J. Biol. Chem.* 275:16408–16413.

56. Ichiyanagi, K., Y. Ishino, …, K. Morikawa. 2000. Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. *J. Mol. Biol.* 300:889–901.

57. Shemella, P. 2008. First principles study of intein reaction mechanisms. PhD thesis, Rensselaer Polytechnic Institute, Troy, NY.

58. Shemella, P. T., and S. K. Nayak. 2010. Identifying the reaction mechanisms of inteins with QM/MM multiscale methods. *In* Computational Modeling in Biomechanics. 469–489.

59. Johansson, D. G. A., G. Wallin, …, T. Härd. 2009. Protein autoproteolysis: conformational strain linked to the rate of peptide cleavage by the pH dependence of the N→O acyl shift reaction. *J. Am. Chem. Soc.* 131:9475–9477.

60. Milner-White, E. J. 1997. The partial charge of the nitrogen atom in peptide bonds. *Protein Sci.* 6:2477–2482.

61. Reference deleted in proof.

62. Chang, J., A. M. Lenhoff, and S. I. Sandler. 2007. Solvation free energy of amino acids and side-chain analogues. *J. Phys. Chem. B.* 111:2098–2106.

63. Pereira, B., S. Jain, and S. Garde. 2006. Quantifying the protein core flexibility through analysis of cavity formation. *J. Chem. Phys.* 124:74704.

64. Reference deleted in proof.

65. Reed, A. E., L. A. Curtiss, and F. Weinhold. 1988. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* 88:899–926.

66. Wu, W. Y., A. R. Gillies, …, D. W. Wood. 2010. Self-cleaving purification tags re-engineered for rapid Topo® cloning. *Biotechnol. Prog.* 26:1205–1212.