

Autocatalytic Intramolecular Isopeptide Bond Formation in Gram-Positive Bacterial Pili: A QM/MM Simulation

Xiangqian Hu,[†] Hao Hu,[‡] Jeffrey A. Melvin,[§] Kathleen W. Clancy,[†]
Dewey G. McCafferty,^{*,†} and Weitao Yang^{*,†}

Department of Chemistry, Duke University, Durham, North Carolina 27708; Department of Chemistry, The University of Hong Kong; and Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27708, United States of America

Received August 19, 2010; E-mail: dewey.mccafferty@duke.edu; weitao.yang@duke.edu

Abstract: Many Gram-positive pathogens possess external pili or fimbriae with which they adhere to host cells during the infection process. Unusual dual intramolecular isopeptide bonds between Asn and Lys side chains within the N-terminal and C-terminal domains of the pilus subunits have been observed initially in the *Streptococcus pyogenes* pilin subunit Spy0128 and subsequently in GBS52 from *Streptococcus agalactiae*, in the BcpA major pilin of *Bacillus cereus* and in the RrgB pilin of *Streptococcus pneumoniae*, among others. Within each pilin subunit, intramolecular isopeptide bonds serve to stabilize the protein. These bonds provide a means to withstand large external mechanical forces, as well as possibly assisting in supporting a conformation favored for pilin subunit polymerization via sortase transpeptidases. Genome-wide analyses of pili-containing Gram-positive bacteria are known or suspected to contain isopeptide bonds in pilin subunits. For the autocatalytic formation of isopeptide cross-links, a conservation of three amino acids including Asn, Lys, and a catalytically important acidic Glu (or Asp) residue are responsible. However, the chemical mechanism of how isopeptide bonds form within pilin remains poorly understood. Although it is possible that several mechanistic paths could lead to isopeptide bond formation in pili, the requirement of a conserved glutamate and highly organized positioning of residues within the hydrophobic environment of the active site were found in numerous pilin crystal structures such as Spy0128 and RrgB. This suggests a mechanism involving direct coupling of lysine side chain amine to the asparagine carboxamide mediated by critical acid/base or hydrogen bonding interactions with the catalytic glutamate residue. From this mechanistic perspective, we used the QM/MM minimum free-energy path method to examine the reaction details of forming the isopeptide bonds with Spy0128 as a model pilin, specifically focusing on the role of the glutamate in catalysis. It was determined that the reaction mechanism likely consists of two major steps: the nucleophilic attack on C γ by nitrogen in the unprotonated Lys ϵ -amino group and, then two concerted proton transfers occur during the formation of the intramolecular isopeptide bond to subsequently release ammonia. More importantly, within the dual active sites of Spy0128, Glu¹¹⁷, and Glu²⁵⁸ residues function as crucial catalysts for each isopeptide bond formation, respectively, by relaying two proton transfers. This work also suggests that domain–domain interactions within Spy0128 may modulate the reactivity of residues within each active site. Our results may hopefully shed light on the molecular mechanisms of pilin biogenesis in Gram-positive bacteria.

Introduction

Pili are hairlike polymeric proteins that play important roles in the biological functions of many bacteria. These filamentous, multisubunit pili reside extracellularly and can be involved in the transfer of genetic material, induction of signaling in host cells, twitching motility, and other critical aspects of colonization.¹ Pili of pathogenic bacteria are often also major virulence factors and important vaccine candidates. As such, studying the structure and the formation of the pilus is vital for gaining an understanding of pilus biogenesis and may influence the discovery of new drugs or suitable vaccines in the fight against virulent bacteria. For Gram-negative organisms, the structures

and functions of pili have been studied extensively. For instance, type I and type IV pili, which are long (1 to 4 μm), thin (5 to 8 nm), and flexible, have been characterized well in terms of their structures and assembly mechanisms.^{2–7} However, the covalent and three-dimensional structure of Gram-positive pili and the mechanism of their assembly have remained a mystery

- (2) Choudhury, D.; Thompson, A.; Stojanoff, V.; Langermann, S.; Pinkner, J.; Hultgren, S. J.; Knight, S. D. *Science* **1999**, 285, 1061.
- (3) Craig, L.; Pique, M. E.; Tainer, J. A. *Nat. Rev. Microbiol.* **2004**, 2, 363.
- (4) Craig, L.; Volkmann, N.; Arvai, A. S.; Pique, M. E.; Yeager, M.; Egelman, E. H.; Tainer, J. A. *Mol. Cell.* **2006**, 23, 651.
- (5) Hahn, E.; Wild, P.; Hermanns, U.; Sebbel, P.; Glockshuber, R.; Haner, M.; Taschner, N.; Burkhard, P.; Aebi, U.; Muller, S. A. *J. Mol. Biol.* **2002**, 323, 845.
- (6) Parge, H. E.; Forest, K. T.; Hickey, M. J.; Christensen, D. A.; Getzoff, E. D.; Tainer, J. A. *Nature* **1995**, 378, 32.
- (7) Remaut, H.; Waksman, G. *Curr. Opin. Struct. Biol.* **2004**, 14, 161.

[†] Department of Chemistry, Duke University.

[‡] Department of Chemistry, The University of Hong Kong.

[§] Department of Biochemistry, Duke University Medical Center.

(1) Yeates, T. O.; Clubb, R. T. *Science* **2007**, 318, 1558.

until recently.^{1,8–15} Compared to Gram-negative pili, the Gram-positive pili assembled by sortases (bacterially encoded transpeptidase enzymes) are extremely thin (2 to 3 nm) because the Gram-positive pilin polymer chain is composed of many covalently bonded copies of a *single* backbone pilus. Gram-positive bacteria employ intramolecular isopeptide bonds to stabilize pili folded structures that are displayed extracellularly, as an alternative to disulfide bonds used to stabilize intracellular protein structures.^{8,16} For example, *Streptococcus pyogenes* infects the human throat and skin cells with adhesive pili, potentially causing necrotizing fasciitis, rheumatic fever, and streptococcal toxic shock syndrome. The structure of the M1 pilin subunit Spy0128 from *S. pyogenes* strain SF370 [also known as group A Streptococcus (GAS)] indicates that the pilin is stabilized by two unique intramolecular isopeptide bonds per subunit.^{8,9} Spy0128 mutants lacking isopeptide bonds exhibit lowered trypsin-resistance and thermal stability.

In Spy0128, one intramolecular isopeptide bond is formed between Lys³⁶ and Asn¹⁶⁸ in the N-terminal domain and the other between Lys¹⁷⁹ and Asn³⁰³ in the C-terminal domain. The most interesting aspect is that both bond formations occur naturally without the aid of any cofactor or enzyme. In the Spy0128 structure, an autocatalytic Glu residue sits proximal to each Lys-Asn isopeptide bond and has been shown by mutational analyses to be essential to generate these bonds.^{8,9} The Glu residue is associated with each bond, forming hydrogen bonds to the isopeptide C=O and NH groups. The hydrogen bonding implies that both glutamic acids, Glu¹¹⁷ and Glu²⁵⁸, are protonated. The packing of the Lys, Asn, and Glu residues within hydrophobic pocket lined with aromatic residues predicts an elevation of the pK_a of the glutamic acid and reduction of the pK_a of the lysine ϵ -amino group. As such, a direct attack mechanism for isopeptide bond formation has been proposed. The protonated Glu polarizes the C=O bond of the Asn side chain and induces positive charge on the amide carbon. Then nucleophilic attack on C γ by nitrogen in the unprotonated Lys ϵ -amino group generates the isopeptide bond. We refer to this mechanism as *inverse protonated*, referring to the state whereby the protonated Glu and unprotonated Lys are catalytically competent protonation states of these side chains.

In this work, we studied the reaction mechanism of the intramolecular isopeptide bond formations using ab initio QM/MM simulations and revealed the reaction mechanism with atomic detail. We suggest that if indeed the direct attack mechanism occurs as hypothesized, the reaction must occur in the *inverse protonated* state. Then, the nucleophilic attack on C γ in Asn by nitrogen in the Lys ϵ -amino group forms the isopeptide bond. Our simulations also demonstrated that concerted transfers of two protons during the formation of the intramolecular isopeptide bond are vital to release ammonia and fulfill the entire reaction. Moreover, this reaction is catalyzed

by the key Glu residue functioning as a proton relay medium during proton transfers. We also conducted comparative reaction mechanism studies on both locations (Lys³⁶-Asn¹⁶⁸ and Lys¹⁷⁹-Asn³⁰³) as they formed isopeptide bonds, respectively. The computed reaction barriers for both isopeptide bonds suggest that the isopeptide bond (Lys³⁶-Asn¹⁶⁸) in the N-terminal domain may be formed prior or simultaneously to the other isopeptide bond (Lys¹⁷⁹ and Asn³⁰³) in the C-terminal domain. This is shown to be an effect of the protein domain–domain interaction.

Methods

Computational Methods for QM/MM Simulations. In this work, all of the QM/MM simulations employed the recently developed QM/MM-minimum free energy path (MFEP) method. The details of the QM/MM-MFEP approach have been discussed in our previous publications.^{17–19} The key ingredient of the QM/MM-MFEP is that the geometry of the QM subsystem is optimized on the potential of mean force (PMF) surface of the subsystem conformations.^{17–19} For the MM energy $E_{\text{MM}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}})$, all of the MM atoms are described by the CHARMM force field²⁰ and TIP3P water model.²¹ All of the QM energy calculations were performed at the B3LYP/6-31+G*^{22–25} level using the locally modified Gaussian 03 program.²⁶ The QM/MM interaction energy $E_{\text{QM/MM}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}})$ includes the classical point charge and Lennard–Jones interactions between the QM and MM subsystems. The QM charges are fitted from the electrostatic potential (ESP) of the frozen QM geometry^{17,18,27} and are used to calculate the point charge interactions in $E_{\text{QM/MM}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}})$ without consideration of the charge polarization effect.²⁸ The covalent interactions between the QM and MM subsystems are characterized by the pseudobond approach.²⁹ In this work, a new set of pseudobond parameters, fitted with amino acids as the training set, was used.³⁰

The reaction path on the PMF surface can be readily optimized with the QM/MM-MFEP method. Both the nudged-elastic-band (NEB)³¹ and quadratic string methods (QSM)^{32,33} were used here. In both algorithms, the full QM degrees of freedom are used to construct a discrete reaction path from the reactant state to the product state. The QM/MM-MFEP method can be combined with NEB or QSM smoothly. The optimization procedure^{17,18} is the following: (i) the discrete reaction path is optimized using NEB or QSM in a fixed and finite MM ensemble; (ii) the new MM ensemble is generated by the MD simulation with the fixed QM conformation; (iii) go to step (i) until the convergence criteria on the PMF and path differences and the normalized gradients of the entire path are satisfied. The iterative sampling and optimization algorithm

(8) Kang, H. J.; Baker, E. N. *J. Biol. Chem.* **2009**, *284*, 20729.

(9) Kang, H. J.; Coulbaly, F.; Clow, F.; Proft, T.; Baker, E. N. *Science* **2007**, *318*, 1625.

(10) Scott, J. R.; Zahner, D. *Mol. Microbiol.* **2006**, *62*, 320.

(11) Telford, J. L.; Barocchi, M. A.; Margarit, I.; Rappuoli, R.; Grandi, G. *Nat. Rev. Microbiol.* **2006**, *4*, 509.

(12) Ton-That, H.; Marraffini, L. A.; Schneewind, O. *Biochim. Biophys. Acta, Mol. Cell. Res.* **2004**, *1694*, 269.

(13) Ton-That, H.; Mazmanian, S. K.; Alksne, L.; Schneewind, O. *J. Biol. Chem.* **2002**, *277*, 7447.

(14) Ton-That, H.; Schneewind, O. *Mol. Microbiol.* **2003**, *50*, 1429.

(15) Ton-That, H.; Schneewind, O. *Trends Microbiol.* **2004**, *12*, 228.

(16) Alegre-Cebollada, J.; Badilla, C. L.; Fernandez, J. M. *J. Biol. Chem.* **2010**, *285*, 11235.

(17) Hu, H.; Lu, Z. Y.; Parks, J. M.; Burger, S. K.; Yang, W. T. *J. Chem. Phys.* **2008**, *128*, 18.

(18) Hu, H.; Lu, Z. Y.; Yang, W. T. *J. Chem. Theory Comput.* **2007**, *3*, 390.

(19) Hu, H.; Yang, W. T. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573.

(20) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.

(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(22) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(23) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(24) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(25) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(26) Frisch, M. J. In *Gaussian 03*; C.02 ed.; Gaussian, Inc.: Wallingford CT, 2004.

(27) Hu, H.; Lu, Z. Y.; Yang, W. T. *J. Chem. Theory Comput.* **2007**, *3*, 1004.

(28) Lu, Z. Y.; Yang, W. T. *J. Chem. Phys.* **2004**, *121*, 89.

(29) Zhang, Y. K.; Lee, T. S.; Yang, W. T. *J. Chem. Phys.* **1999**, *110*, 46.

(30) Parks, J. M.; Hu, H.; Cohen, A. J.; Yang, W. T. *J. Chem. Phys.* **2008**, *129*, 6.

(31) Jónsson, H.; Mills, G.; Jacobsen, K. W. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific: Singapore, 1998, p 385.

(32) Burger, S. K.; Yang, W. T. *J. Chem. Phys.* **2007**, *127*, 7.

(33) Burger, S. K.; Yang, W. T. *J. Chem. Phys.* **2006**, *124*, 13.

reduces the computational cost of QM/MM simulations dramatically with high accuracy and this scheme has been successfully applied to study the reaction mechanisms of several enzymes.^{34,35} Moreover, the QM/MM-MFEP approach was further extended to calculate solution redox free energies for metal complexes and organic molecules.³⁶

Computational Construction of the Initial Reactant Structure. The *S. pyogenes* Spy0128 crystal structure from the protein data bank (PDB ID: 3B2M) at 2.2 Å resolution is shown in Figure S1 of the SI and was used to build the initial reactant structure.^{8,9} This protein includes 340 amino acid residues, but the first 16 residues were removed in our work since they are disordered in the structure and do not encompass the isopeptide bond regions. The pilin crystal structure has an elongated two-domain structure (i.e., N-terminal and C-terminal domains). Each domain has one unique intramolecular isopeptide bond between Lys³⁶ and Asn¹⁶⁸ in the N-terminal domain and between Lys¹⁷⁹ and Asn³⁰³ in the C-terminal domain. The N–C isopeptide bond linking the Lys ε-amino group and Asn for each domain was broken to construct the initial state of the protein. The missing NH₂ group was added back to Asn¹⁶⁸ and Asn³⁰³ for each isopeptide bond with the program Molden.³⁷ All of the hydrogen atoms were added using the web service MolProbity.³⁸ The entire protein was solvated in a rectangular box of size 66 × 84 × 130 Å³. The total number of water molecules is 20,820, including 136 crystal water molecules.

MD Simulations of Pilus Proteins. The geometry of the entire system was first optimized with MM force fields, and then the water molecules were warmed up gradually to 300 K. The entire system was warmed up through an 800 ps molecular dynamics (MD) simulation in which the atoms at the protein backbone were restrained by a harmonic force of 40 kcal/mol/Å² and then subjected to a 320 ps MD simulation using the same restraint with a harmonic force of 20 kcal/mol/Å². The multiple-time step algorithm³⁹ was used, in which the integration step sizes were 2 fs for short-range forces, 4 fs for medium-range forces, and 8 fs for long-range electrostatic forces. The PME method was used for computing the long-range electrostatic interactions.⁴⁰ All bonds of water molecules were constrained by the SHAKE algorithm.⁴¹ An 8 and 12 Å dual cutoff was employed to generate the nonbonded pair lists, which were updated every 32 fs. The temperature and pressure of the system were maintained at 300 K and 1 bar with the Berendsen thermostat and manostat.⁴²

QM/MM-MFEP Simulations of the Isopeptide Bond Formation. X-ray crystallographic and mutagenesis studies^{8,9} show that Glu¹¹⁷ and Glu²⁵⁸ play vital roles in both isopeptide bond formations. Two active sites including Lys³⁶, Asn¹⁶⁸, and Glu¹¹⁷ as the QM subsystem A (or Active Site A), and Lys¹⁷⁹, Asn³⁰³, and Glu²⁵⁸ as the QM subsystem B (or Active Site B), were studied respectively. (See Figure S1 of the SI for details.) All of the C_α atoms on the side chains as the boundary atoms between the QM and MM subsystems were modeled by the pseudobond approach^{29,30} with the parameters in ref 30. The total number of atoms in each QM subsystem is 36. For single-point geometry optimizations (for

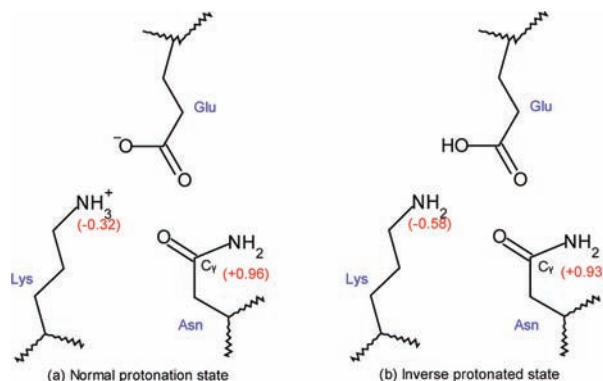


Figure 1. Two possible protonation states of the active site for isopeptide bond formation. All three residue names are labeled in blue. The ESP charges on the nitrogen atom in the Lys epsilon amino group and the C_γ atom in the Asn group are shown in red for both protonation states.

instance, the reactant geometry and the discrete geometries generated by coordinate driving approach to construct the initial reaction path), the modified “Gau_external” script in Gaussian 03²⁶ combined with the QM/MM-MFEP method was used. 80 ps of the MD sampling was used to calculate the QM free energy perturbations and free energy gradients (see eqs 1 and 2 in ref 17). For the reaction path optimization, the coordinate driving approach was used to generate the initial reaction path. 40 ps of the MD sampling with NEB was performed to optimize the path, and then 80 ps was used in the late stage. 160 ps of the MD sampling was also performed to verify that 80 ps was sufficient to converge the reaction path optimizations. A dual cutoff of 9 and 15 Å was used for the QM/MM-MFEP calculations.¹⁷ The SHAKE algorithm was only used on water molecules. The integration time steps were 1 fs for short-range forces, 4 fs for medium-range forces, and 8 fs for long-range electrostatic forces.

Results and Discussion

Normal versus Inverse Protonated States. Two different protonation states of the side chains of Lys and Glu in the active sites A and B from Spy0128 are shown in Figure 1: the normal zwitterionic protonation state, in which Lys is protonated (with one positive charge) and Glu is deprotonated (with one negative charge), and the *inverse protonated* state, in which both Lys and Glu side chains are charge neutral. We first simulated the proton transfer reaction from normal to *inverse protonated* state for Active Site A. The initial geometry was obtained from the last snapshot of the 640 ps MD simulation without any restrained forces. Then the geometry of Active Site A in the normal protonation state was optimized by QM/MM-MFEP as shown in Figure S7 of the SI, and the distance difference ($d_0 - d_1$) shown in Figure S7 of the SI) between H–N in Lys³⁶ and H–O in Glu¹¹⁷ is used to drive the protonation state from the normal to the *inverse protonated* one. The profile of potential of mean force (PMF) during this proton transfer was plotted in Figure 2. The free energy difference between the normal and inverse states is 5.4 kcal/mol and the transition barrier is only 7.5 kcal/mol for Active Site A. Note that the optimized bond distances of H–N in Lys and H–O in Glu for both active sites do not vary with different initial geometry preparation or the choice of the active site. This suggests that the proton-transfer step between two different protonation states is not the rate limiting step for both active sites.

We further optimized the geometry of Active Site A in the gas phase and aqueous solution (simulated by polarizable continuum models) in the absence of surrounding protein structure using Gaussian 03.²⁶ Our results show that the normal

(34) Hu, H.; Boone, A.; Yang, W. T. *J. Am. Chem. Soc.* **2008**, *130*, 14493.

(35) Parks, J. M.; Hu, H.; Rudolph, J.; Yang, W. T. *J. Phys. Chem. B* **2009**, *113*, 5217.

(36) Zeng, X. C.; Hu, H.; Hu, X. Q.; Yang, W. T. *J. Chem. Phys.* **2009**, *130*, 8.

(37) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123.

(38) Lovell, S. C.; Davis, I. W.; Adrendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 437.

(39) Humphreys, D. D.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem.* **1994**, *98*, 6885.

(40) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(41) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(42) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

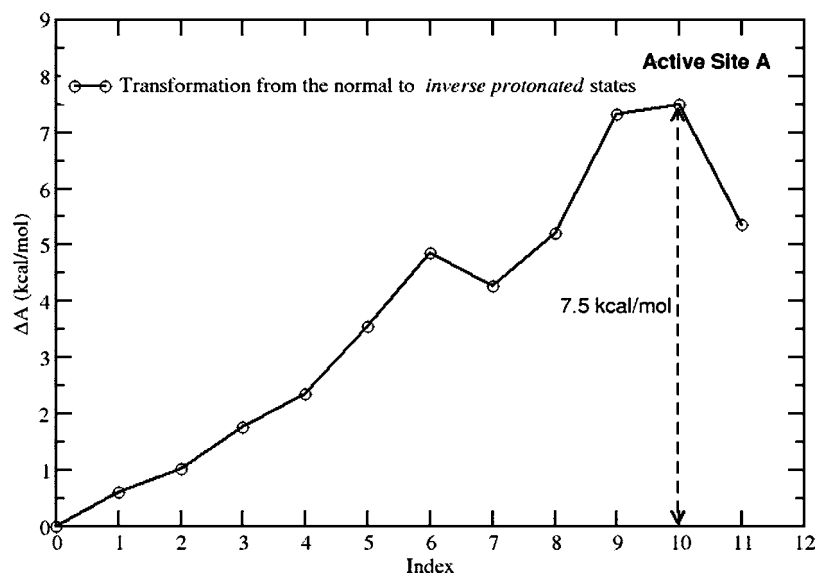


Figure 2. Potential of mean force of the transformation from normal to *inverse protonated* states for Active Site A.

protonation state is only stable in aqueous solution, while the *inverse protonated* state is very stable in gas phase instead. Within the Spy0128 structure, we found that many aromatic amino acids surround both active sites (see Figure S2 of the SI for details). This hydrophobic protein environment can thus stabilize the *inverse protonated* state similar to the situation in the gas phase. Moreover, when the active site is in the *inverse protonated* state, as shown in Figure 1, the ESP charge of the nitrogen atom in the Lys ϵ -amino group becomes partially negatively charged to -0.58 as compared to -0.32 in the normal protonated state. The more negatively charged nitrogen atom facilitates the next reaction step, i.e., the nucleophilic attack on C_γ of the Asn group. We also demonstrated that the direct nucleophilic attack in a normally zwitterionic state seems unlikely because the calculated transition barrier is too high (>35 kcal/mol in Figure S3 of the SI) as opposed to the mechanistic model involving the *inverse protonated* reactant state. Therefore, we conclude that the isopeptide bond formation within a direct acylation mechanism catalyzed by the proximal glutamates likely occurs in the *inverse protonated* state for both active sites. This result is consistent with the postulated mechanism on experimental mutation studies of the glutamates.^{8,9}

Reaction Mechanism. The active site structures of pilins without the isopeptide bonds are not maintained well. In the absence of the isopeptide cross-link, the structures show large conformational fluctuations during simulations compared to the crystal structure of the mature cross-linked Spy0128 pilin (see Figure S5 of the SI for details). Note that these large fluctuations support the view that the isopeptide bonds are critical to stabilize the global protein structure.^{8,9} As such, the initial reactant geometry of Active Site B was generated from the 320 ps MD simulation with restraints (i.e., a harmonic force of 20 kcal/mol/Å²) on the heavy atoms of protein backbone in order to study the subtle reaction process of the isopeptide bond formations. This geometry was further optimized by the QM/MM-MFEP method in the *inverse protonated* state without any restraints and is shown in Figure 3a along with the driving coordinates d_{CN} for the nucleophilic attack step and d'_{CN} for the ammonia formation.

The optimized reaction path of the isopeptide bond formation for Active Site B is shown in Figure 4. The reaction mechanism in Figure 5 was revealed for Active Site B: (1) the activation

barrier of the nucleophilic attack step is low (10.2 kcal/mol) as expected, and (2) the ammonia formation step can occur with a low energy barrier (14.7 kcal/mol). At the nucleophilic attack step, the unprotonated Lys¹⁷⁹ side chain ϵ -amino group rotates to attack the C_γ of Asn³⁰³ due to the strong electrostatic attraction.^{8,9} The second step is that the protonated Glu²⁵⁸ transfers its proton to the amino group of Asn³⁰³. Meanwhile, another proton in the Lys¹⁷⁹ ϵ -amino group is transferred to Glu²⁵⁸. The geometry of the transition state for the inter-residue proton transfer step is depicted in Figure 3b. For this transition state, one proton is transferred to the amino group of Asn³⁰³. The ammonia is formed, and it is ready to exit the active site (the bond distance between C and N in Asn³⁰³ becomes 1.69 Å). The other proton forms a weakened O–H bond with Glu²⁵⁸ (the bond distance between O and H in Glu²⁵⁸ is 1.25 Å). The long distance (1.54 Å) of the isopeptide bond indicates that the bond is still weak. As shown in Figure 3c for the product state, the isopeptide bond becomes much stronger (the bond distance is 1.34 Å) and the ammonia remains (the distance between C and NH₃ is 2.76 Å). In addition, the product geometry is very similar to the original crystal structure (illustrated by the black line in Figure 3c). The major geometry difference is the position of the carboxyl group in Glu²⁵⁸, which is influenced by the presence of ammonia in the product state of our model. This suggests that the optimized reaction path can produce the correct product geometry.

We further studied Active Site A using the last snapshot of the above 320 ps MD simulation with the restraints on the protein backbone. (Note that another reaction mechanism shown in Figure S6 of the SI was also studied and more details are discussed in SI-2 and in Figures S7 and S8 of the SI. This reaction mechanism is not supported in this work because of the high energy barrier obtained.) The reactant geometry was then optimized by the QM/MM-MFEP method in the *inverse protonated* state and is shown in Figure 6a. The reactant geometry of Active Site A includes two hydrogen bonds between Glu¹¹⁷ and Asn¹⁶⁸. Compared to the reactant geometry of Active Site B with only one weak hydrogen bond shown in Figure 3a, the carboxyl group of Glu¹¹⁷ of Active Site A is oriented differently. Following the same driving schemes for Active Site B to obtain the initial reaction path, the reaction path of the isopeptide bond formation for Active Site A is

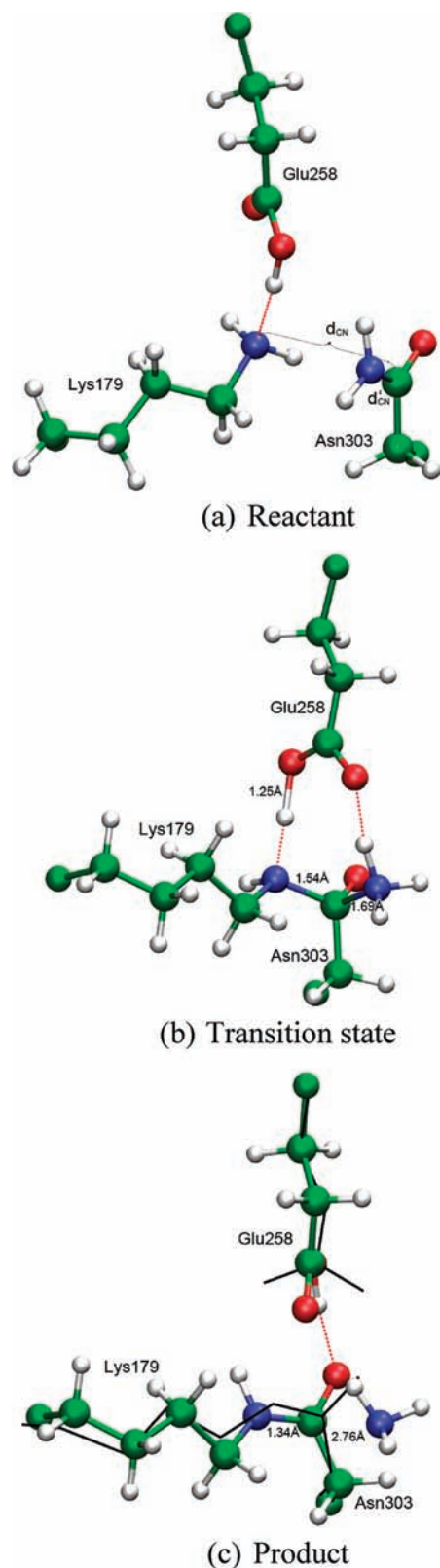


Figure 3. (a) The reactant geometry of Active Site B optimized by the QM/MM-MFEP method. (b) The geometry details of the transition state for two interproton transfers of Active Site B. (c) The geometry difference between the final product state (shown in the stick-and-ball style) and the original crystal structure of the pilus (shown in the black line) for Active Site B. Carbon atoms are in green; hydrogen atoms are in white; nitrogen atoms are in blue; and the oxygen atoms are in red. The red dashed lines represent the hydrogen bonds between oxygen and hydrogen atoms. All other figures follow the same representations.

optimized and shown in Figure 7. Overall, the final isopeptide bonds and their formation through nucleophilic attack are very similar for the active sites A and B. However, the activation barrier of the second step for the inter-residue proton transfer from Glu¹¹⁷ to Asn¹⁶⁸ in Active Site A (31.4 kcal/mol) is much higher than in Active Site B (14.7 kcal/mol). We showed in the following section that this barrier difference between active sites A and B is caused by the reactant geometry difference between active sites A and B arising from the different locations of active sites in the pilus (see Figure S1 of the SI).

After submission of this manuscript for review, Schwarz-Linek and co-workers⁵⁰ report in a communication a biophysical and computational analysis of the spontaneously formed Lys-Asp isopeptide bond in CnaB2 of FbaB of invasive *S. pyogenes* strains. They concluded that isopeptide bonds in this protein increase the folded stability and through simulation have provided some insight into a possible mechanism of Lys-Asp isopeptide formation that can be related in a supportive manner to this study. Although not a pilin and composed of a Lys-Asp versus Lys-Asn cross-link, a similar reaction mechanism is supported: the Lys-Asp isopeptide bond formation is predicted to occur with side chains in neutral protonation states, and two proton transfers are required to release one water molecule through Glu near the active site. However, a distinguishing feature of the Spy0128 system is that the two Lys-Asn isopeptide bonds in two domains of Spy0128 pilin participate in domain–domain interactions, influencing the Lys-Asn isopeptide bond formations as discussed in the following session. Nonetheless, elements of the isopeptidation mechanism proposed for Spy0128 may be potentially echoed in other systems such as within the CnaB2 domain of FbaB.

Timing of Isopeptide Bond Formation. For Active Site A, although different reactant geometries and reaction schemes were tested, none of them appear to be plausible because of high activation barriers (see Figures 7 and S8 of the SI). The reason resides in the specific protein environments of Active Site A. According to the pilin crystal structure shown in Figure S1 of the SI, Active Site B is at the end of the C-terminal domain and Active Site A is located at the interface between the N-terminal and C-terminal domains. We hypothesized that this protein domain–domain interaction may significantly influence the geometry of Active Site A and thus the reaction barriers of the isopeptide bond formation. To verify this, we removed the C-terminal domain and used only the N-terminal domain of the pilus (which was cut out from residues 30 to 171 in the crystal structure, named Pilus_Cut) in the QM/MM simulations of isopeptide bond formation for Active Site A (see Figure S4 of the SI).

Denoted Pilus_Cut, this structure only includes the N-terminal domain and contains only Active Site A. The initial geometry of Active Site A was obtained from the 320 ps MD simulation with the restraints on the protein backbone. The reactant geometry was then optimized by the QM/MM-MFEP method in the *inverse protonated* state and is shown in Figure 6b. Indeed, compared to the superimposed reactant geometry of Active Site B shown in the purple line of Figure 6b, the backbone of Active Site A in the Pilus_Cut structure without the C-terminal domain becomes very similar to that of Active Site B. This suggests that the excised C-terminal domain can definitely influence the structure of Active Site A. Compared to the reactant geometry of Active Site A of the wild-type pilin in Figure 6a, Active Site A of Pilus_Cut only forms one hydrogen bond with the less ordered structure. In contrast, when the C-terminal domain is present in the wild-type pilus, Active

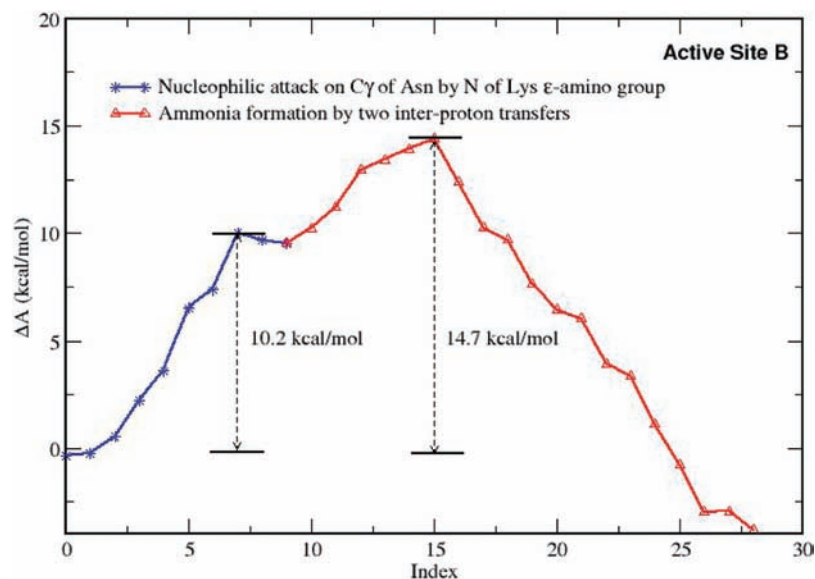


Figure 4. Potential of mean force of the isopeptide bond formation for Active Site B.

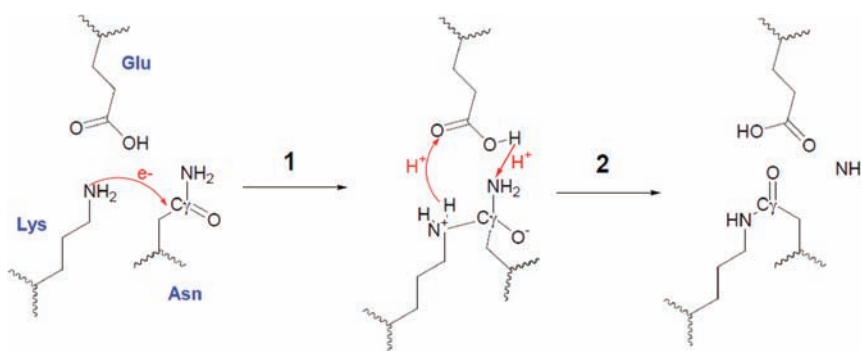


Figure 5. Putative direct nucleophilic attack reaction mechanism for the isopeptide bond formation in a pilin subunit. The first step is nucleophilic attack on C γ in the Asn group by nitrogen in the unprotonated Lys ϵ -amino group. The second step involves two concerted proton transfers to release the ammonia.

Site A is apt to form hydrogen bond networks that order the structure (see Figures 6a and S7 of the SI). Hence, the domain–domain interactions in the pilus can affect the reactant geometry of Active Site A dramatically.

To examine the effects of domain–domain interaction on the reaction barriers, we optimized the reaction path of the isopeptide bond formation for Active Site A of Pilus_Cut using the previous coordinate driving procedure. As shown in Figure 8, the nucleophilic attack becomes the rate limiting transition of Active Site A for Pilus_Cut with the barrier lowered to 24.5 kcal/mol significantly compared with 31.4 kcal/mol in Figure 7 for Active Site A of the wild-type pilin. Although the total activation barrier for Active Site A in Pilus_Cut still appears high compared with that for Active Site B in the wild-type pilus, Pilus_Cut is only one model for the N-terminal domain to show how the domain–domain interactions influence the reaction barrier in pilins. The N-terminal domain may fold to another state during the assembly of a pilus in order to carry out the reaction. Particularly, for Pilus_Cut, the carboxyl group of Glu¹¹⁷ and amino groups of Asn¹⁶⁸ and Lys³⁶ need to rotate the hydrogen atoms into the appropriate positions in order to allow the nucleophilic attack on C γ of Asn¹⁶⁸ by the nitrogen of Lys³⁶. These extra rotations increase the activation barrier of the nucleophilic attack step more than 10 kcal/mol compared to that for Active Site B. However, the relative activation barrier for the second reaction step in Active Site A for Pilus_Cut to

generate the ammonia is lowered significantly. As shown in Figure 8, the relative barrier to the intermediate state after the nucleophilic attack step is only 5.4 kcal/mol, compared to 19.8 kcal/mol of Active Site A for the wild-type pilin in Figure 7. This proton transfer step still involves two inter-residue proton transfers same as Active Site B for the normal pilin. In short, by removing the C-terminal domain, the overall activation barrier of Active Site A drops by 7 kcal/mol. This suggests that the domain–domain interactions between N- and C-terminal domains can significantly influence the isopeptide bond formation at Active Site A.

Isopeptide bonds occur in many proteins, but most are formed intermolecularly with the assistance of enzyme catalysts acting on nascent folded polypeptides. Such is seen in the case of ubiquitination of proteins by ubiquitin lyases⁴³ and transglutamination protein cross-linking via transglutaminases.^{44,45} *Autocatalytic* isopeptidation has been observed in relatively few proteins outside pilin subunits, such as in the bacteriophage HK97,^{46,47} where capsid subunits are covalently cross-linked

(43) Pickart, C. M. *Annu. Rev. Biochem.* **2001**, *70*, 503.

(44) Greenberg, C. S.; Birckbichler, P. J.; Rice, R. H. *FASEB J.* **1991**, *5*, 3071.

(45) Weisel, J. W. In *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers*; 200570, p 247.

(46) Wikoff, W. R.; Liljas, L.; Duda, R. L.; Tsuruta, H.; Hendrix, R. W.; Johnson, J. E. *Science* **2000**, *289*, 2129.

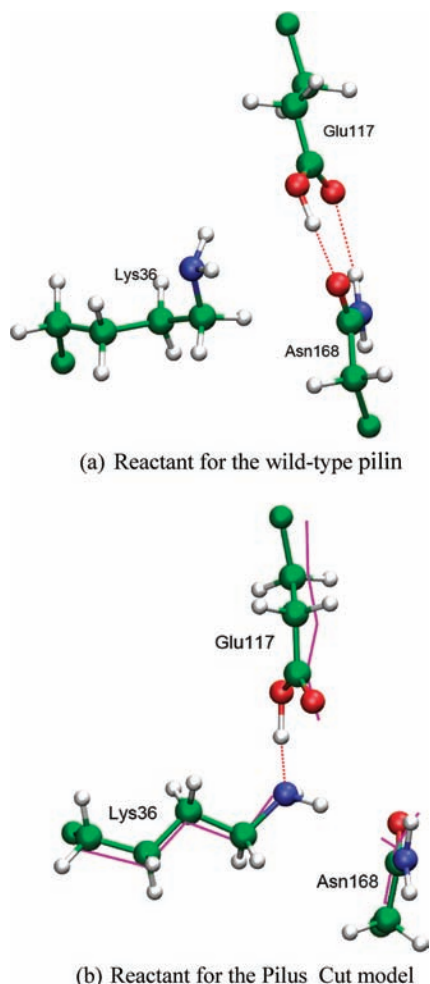


Figure 6. (a) The reactant geometry of Active Site A optimized by the QM/MM-MFEP method in the *inverse protonated* state. (b) The reactant geometry of Active Site A optimized by the QM/MM-MFEP method in the *inverse protonated* state of Pilus_Cut. The initial geometry was generated from the restrained MD simulations. The reactant geometry of Active Site B (shown in Figure 3a) was aligned to Active Site A and its backbone is shown in purple.

between Lys and Asn residues to form locked intermolecular catenated rings. Similar to Spy0128 and other pilins, HK97 capsid shares a similar requirement for a Glu residue to facilitate isopeptide bond formation. However, HK97 capsid cross-linking is initiated post-assembly into the pro-capsid particle and can be triggered by exposure to 6 M urea. In contrast, Spy0128 expressed in the heterologous host organism *E. coli* are isolated fully cross-linked. This suggests that isopeptide bonds form soon after protein expression and not due to recruitment to the inner membrane or encounter with late stage pilin polymerization enzymatic machinery, including the sortase transpeptidase responsible for pilin polymerization or the housekeeping SrtA sortase responsible for covalent attachment of pili to the peptidoglycan precursor Lipid II. Given these observations and results from the QM/MM simulations, it is intriguing to speculate that isopeptide bond formation in Spy0128 as well as other pilins, may occur fairly early in the life of the protein. Within the autocatalytic direct nucleophilic attack mechanism, barrier comparison between two active sites strongly suggests

that the isopeptide bond at Active Site A in the N-terminal domain may be formed earlier than, or simultaneously with, the other isopeptide bond at Active Site B in the C-terminal domain *because the domain–domain interactions among one pilus subunit have a significant adverse impact on the isopeptide bond formation of Active Site A in the N-terminal domain*. Similarly, domain–domain interactions might further influence the rate of amidation and thus translate into timing differences for cross-linking. These observations could support a modular cross-linking reaction from N to C terminal domains as the protein folds and exits the ribosome machinery. Alternatively, the protein may adopt a conformation that becomes optimal for autocatalytic amidation as it encounters factors that may alter global or local protein structure within the intracellular milieu. Nonetheless, both isopeptide bonds are likely formed prior to being polymerized into mature pilus fibers by the bacteria.

Are the mechanisms of autocatalytic isopeptide bond formation generalized among proteins that exhibit a requirement for a catalytically essential glutamate? There are multiple mechanisms one can evoke to create an isopeptide connection within a protein. In this work, we have compared two plausible direct nucleophilic attack mechanisms assisted by glutamate protonation and suggest that within these mechanistic postulates, the inverse protonation pathway is markedly energetically favored. A direct attack mechanism is also supported by inspection of the highly organized structure of the hydrophobic active site and, in Spy0128, mutagenesis studies that validate an essential role for Glu in catalyzing isopeptide bond formation with their respective Lys/Asn pairs. Bacteriophage HK97 capsid assembly is similar to pilin cross-linking in that it requires the presence of a similar ensemble of residues positioned closely within a defined hydrophobic active site.⁴⁶ This hydrophobic environment likely alters the pK_a and nucleophilicity of the Lys and Asn residues in HK97 as we predict occurs in Spy0128.

However, it is important to note that at least one additional mechanism⁴⁸ could be evoked to produce isopeptide bonds in pilus subunits, while appearing to be autocatalytic based on analysis of the reaction products. Asparagine side chains within peptides and proteins can form cyclic succinimide intermediates (shown in Figure S9 of the SI) by condensation with neighboring residues within the chain.⁴⁹ Typically initiated by conformation or by pH, intramolecular condensation of Asn residues with the peptide backbone yields cyclic succinimide intermediates concomitant with the release of ammonia. Within a hydrophobic environment, an intermediate succinimide might be sufficiently protected from hydrolysis, and available to react with a nucleophilic Lys side chain amine to generate an isoamide link. Within this context, the role of the conserved pilin Glu residues could be to assist in the intermediate succinimide formation or in its breakdown following condensation with a Lys residue. Our group is presently comparatively examining more elaborate mechanistic possibilities using QM/MM-MFEP approaches for comparison to the Glu-catalyzed direct nucleophilic mechanism.

Conclusions

Because the intramolecular isopeptide bonds are formed naturally without any cofactors through an autocatalytic process, its reaction mechanism has precluded detailed mechanistic

(47) Heigstrand, C.; Wikoff, W. R.; Duda, R. L.; Hendrix, R. W.; Johnson, J. E.; Liljas, L. *J. Mol. Biol.* **2003**, *334*, 885.

(48) Zakeri, B.; Howarth, M. *J. Am. Chem. Soc.* **2010**, *132*, 4526.

(49) Lura, R.; Schirch, V. *Biochemistry* **1988**, *27*, 7671.

(50) Hagan, R. M.; Bjornsson, R.; McMahon, S. A.; Schomburg, B.; Braithwaite, V.; Buhl, M.; Naismith, J. H.; Schwarz-Linek, U. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 8421.

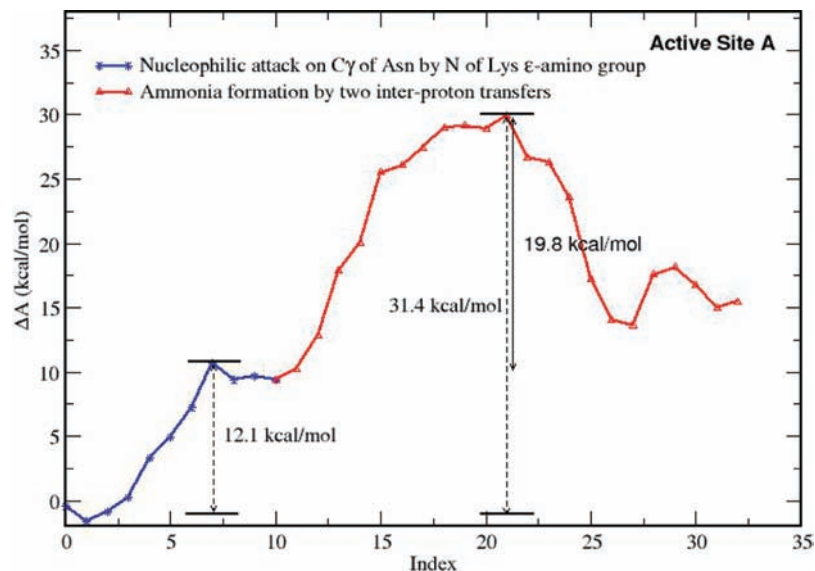


Figure 7. Potential of mean force of the isopeptide bond formation for Active Site A.

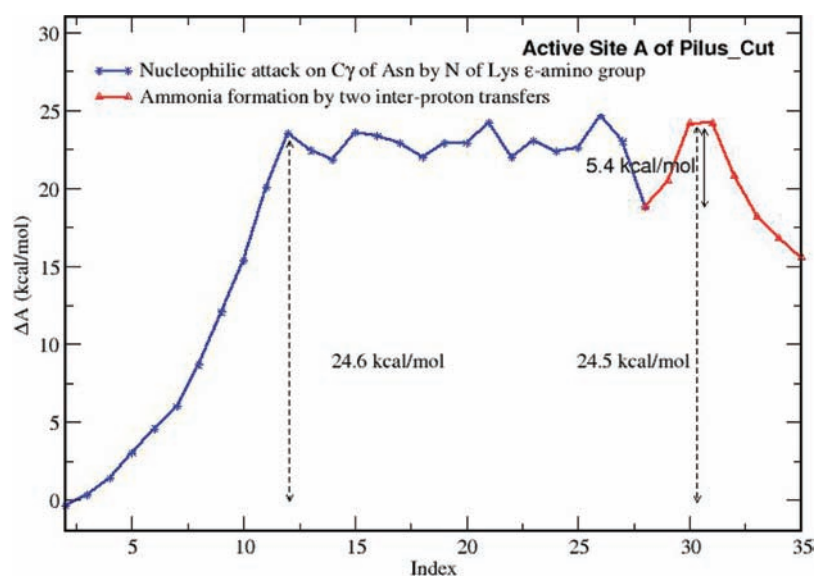


Figure 8. Potential of mean force of the isopeptide bond formation for Active Site A of Pilus_Cut.

investigation from experiments so far. There is no experimental kinetic data yet available. This makes computational studies particularly important. According to our extensive study using the QM/MM-MFEP method, the reaction of the intramolecular isopeptide bond formation must occur in the *inverse protonated* state. The normal and *inverse protonated* states can be switched with low transition barriers. The detailed reaction mechanism of the intramolecular isopeptide bond formation for both active sites in Spy0128 is revealed as following: (i) starting from the *inverse protonated* state, the nitrogen atom with the partial negative charge of the unprotonated Lys attacks C γ with the positive charge of the Asn carbonyl group; (ii) the protonated Glu, as the proton relay medium, rotates and transfers the proton to the leaving amino group of Asn; (iii) meanwhile, the Glu side chain carboxylate accepts the other proton from the Lys ϵ -amino group, generating a neutral amide isopeptide bond; (iv) finally, ammonia is generated and readily eliminated. This reaction mechanism shows that Glu is indispensable to form the isopeptide bonds as the proton relay medium. This conclu-

sion is consistent with the experimental mutation study.⁹ By comparing the activation barriers of two active sites in the pilus, we found that the reaction barrier of Active Site A in the N-terminal domain can be reduced significantly by removing the C-terminal domain. This comparison study suggests that the isopeptide bond at the N-terminal domain may be formed first or simultaneously with the other isopeptide bond at the C-terminal domain.

Acknowledgment. Support from the National Institutes of Health (NIH R01-AI46611 to D.G.M., R01-GM061870 to W.Y.) is greatly appreciated.

Supporting Information Available: Figures S1–S8 and complete refs 20 and 26. All of the Cartesian coordinates of active sites for Figures 3, 6, S5, and S7 optimized by our QM/MM-MFEP method are listed in SI. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA107513T