Published in final edited form as:
J Am Chem Soc. 2011 January 26; 133(3): 478–485. doi:10.1021/ja107513t.

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, NC 27708

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

[††] Department of Biochemistry, Duke University Medical Center, Durham, NC 27708

Abstract

Gram-positive pathogens possess external pili or fimbrae 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 crosslinks, 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 the 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 Cγ 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, Glu^{117} and Glu^{258} residues function as crucial catalysts for each isopeptide bond formation, respectively, by relaying two proton transfers. This study 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.

Corresponding authors: Prof. Dewey G. McCafferty, dewey.mccafferty@duke.edu; Prof. Weitao Yang, weitao.yang@duke.edu.

Supporting Information Available. Several results of other possible reaction path optimizations are available at http://pubs.acs.org. SI includes Figures S1-S8 and complete references 20 and 26. All the Cartesian coordinates of active sites for Figures 3, 6, S5, and S7 optimized by our QM/MM-MFEP method are listed in SI.

Keywords

Pilus; pili; Gram-positive bacteria; isopeptide bond; QM/MM-MFEP; reaction path optimization; QM/MM simulation

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 fighting of virulence bacteria. For Gram-negative organisms, the structures and functions of the pili have been studied extensively. For instance, the 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 its assembly have remained a mystery 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. Furthermore, unlike intracellular proteins that employ disulfides as stabilizing crosslinks, Gram-positive bacteria utilize the intramolecular isopeptide bonds as an alternative way to stabilize pili folded structures when displayed extracellularly.8,16 For instance, *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 Lys^{36} and Asn^{168} in the N-terminal domain and the other between Lys^{179} and Asn^{303} 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, Glu^{117} and Glu^{258}, 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 Cγ 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 atomistic details. 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 Cγ 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 (Lys^{36}-Asn^{168} and Lys^{179}-Asn^{303}) as they formed isopeptide bonds, respectively. The computed reaction barriers for both isopeptide bonds suggest that the isopeptide bond (Lys^{36}-Asn^{168}) in the N-terminal domain may be formed prior or simultaneously to the other isopeptide bond (Lys^{179} and Asn^{303}) 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 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 $E_{MM}(\mathbf{r}_{QM}, \mathbf{r}_{MM})$, 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 $E_{QM/MM}(\mathbf{r}_{QM}, \mathbf{r}_{MM})$ 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 geometry17,18,27 and are used to calculate the point charge interactions in $E_{QM/MM}(\mathbf{r}_{QM}, \mathbf{r}_{MM})$ without consideration of the charge polarization effect.28 The covalent interactions between the QM and MM subsystems are characterized by the pseudo-bond approach.29 In this study, a new set of pseudo-bond 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 procedure17,18 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 reduces the computational cost of QM/MM simulations dramatically with high accuracy and this scheme has been successfully applied to study 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: 2IW5) at 2.2 Å resolution was shown in Figure S1 of 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 Lys^{36} and Asn^{168} in the N-terminal domain and between Lys^{179} and Asn^{303} 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 NH_2 group was added back to Asn^{168} and Asn^{303} 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 $Å^3$. 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/$Å^2$ and then subjected to a 320 ps MD simulation using the same restrain with a harmonic force of 20 kcal/mol/$Å^2$. 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 non-bonded 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

The X-ray crystallographic and mutagenesis studies8,9 show that Glu^{117} and Glu^{258} play vital roles in both isopeptide bond formations. Two active sites including Lys^{36}, Asn^{168}, and Glu^{117} as the QM subsystem A (or Active Site A), and Lys^{179}, Asn^{303}, and Glu^{258} as the QM subsystem B (or Active Site B), were studied respectively. (See Figure S1 in 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 pseudo-bond approach29,30 with the parameters in Ref. 30. The total number of atoms in each QM subsystem is 36. For single-point geometry optimizations (for 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 of 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* sate, in which both Lys and Glu sidechains are charge neutral. We first simulated the proton transfer reaction from normal to *inverse protonated*

states 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 SI and the distance difference ($d_0 - d_1$ shown in Figure S7 of SI) between H-N in Lys^{36} and H-O in Glu^{117} 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 the 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 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 in 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 Cγ 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 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 postulation based on experimental mutation studies of glutamates.[8,9]

Reaction mechanism

The active site structures of pilus 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 in 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/$Å^2$) 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 the low energy barrier (14.7 kcal/mol). At the nucleophilic attack step, the unprotonated Lys^{179} side chain ε-amino group rotates to attack the Cγ of Asn^{303} due to the strong electrostatic attraction. [8,9] The second step is that the protonated Glu^{258} transfers its proton to the amino group of Asn^{303}. Meanwhile, another

proton in the Lys^{179} ε-amino group is transferred to Glu^{258}. 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 Asn^{303}. The ammonia is formed and it is ready to exit the active site (the bond distance between C and N in Asn^{303} becomes 1.69 Å). The other proton forms a weaken O-H bond with Glu^{258} (the bond distance between O and H in Glu^{258} 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 NH_3 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 Glu^{258}, which is influenced by the present 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 restrains on the protein backbone. (Note that another reaction mechanism shown in Figure S6 of SI was also studied and more details are discussed in SI-2 and in Figures S7 and S8 of 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 was shown in Figure 6a. The reactant geometry of Active Site A includes two hydrogen bonds between Glu^{117} and Asn^{168}. Compared to the reactant geometry of Active Site B with only one weak hydrogen bond shown in Figure 3a, the carboxyl group of Glu^{117} 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 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 Glu^{117} to Asn^{168} 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 arisen from the different locations of active sites in the pilus (see Figure S1 in SI).

After submission of this manuscript for review, Schwarz-Linek and coworkers50 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 conclude 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 crosslink, a similar reaction mechanism to what we are proposing was 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 SI). The reason resides in the specific protein environments of Active Site A. According

to the pilus crystal structure shown in Figure S1 of 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 the only 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 in 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 restrains 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 now. 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 pilus 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 Site A is apt to form hydrogen bond networks that order the structure (see Figures 6a and S7 of 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 pilus. Although the total activation barrier for Active Site A in Pilus_Cut still appears high compared with that for Active Site B in pilus, Pilus_Cut is only one model for the N-terminal domain to show how the domain-domain interactions influence the reaction barrier in pilus. The N-terminal domain may fold to another state during the assembling of pilus in order to carry out the reaction. Particularly, for Pilus_Cut, the carboxyl group of Glu^{117}, amino groups of Asn^{168} and Lys^{36} need to rotate the hydrogen atoms into the appropriate positions in order to allow the nucleophilic attack on $C\gamma$ of Asn^{168} by the nitrogen of Lys^{36}. 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 pilus in Figure 7. This proton transfer step still involves two inter-residue proton transfers same as Active Site B for the normal pilus. 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 crosslinking via transglutaminases.44,45 The *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 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 crosslinking is initiated post assembly into the pro-capsid particle and can be triggered by exposure to 6M urea. In contrast, Spy0128 expressed in the heterologous host organism *E. coli* are isolated fully crosslinked. 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 crosslinking reaction from N to C terminal domains as the protein folds and exits the ribosome machinery. Or 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 be polymerized into mature pilin 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 study, 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 crosslinking 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 pKa 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 pilin 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 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.

Conclusion

Because the intramolecular isopeptide bonds are formed naturally without any cofactor through an autocatalytic process, its reaction mechanism has precluded detailed mechanistic 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 that is consistent with experiments. 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 pili is revealed as following: i) starting from the *inverse protonated* state, nitrogen atom with the partial negative charge of the unprotonated Lys attacks Cγ 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, Glu side chain carboxylate accepts the other proton from the Lys ε-amino group, generating a neutral amide isopeptide bond; iv) finally, the ammonia is generated and readily eliminated. This reaction mechanism shows that Glu is indispensible to form the isopeptide bonds as the proton relay medium. This conclusion 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

References

1. Yeates TO, Clubb RT. Science. 2007; 318:1558. [PubMed: 18063774]
2. Choudhury D, Thompson A, Stojanoff V, Langermann S, Pinkner J, Hultgren SJ, Knight SD. Science. 1999; 285:1061. [PubMed: 10446051]
3. Craig L, Pique ME, Tainer JA. Nat Rev Microbiol. 2004; 2:363. [PubMed: 15100690]
4. Craig L, Volkmann N, Arvai AS, Pique ME, Yeager M, Egelman EH, Tainer JA. Mol Cell. 2006; 23:651. [PubMed: 16949362]
5. Hahn E, Wild P, Hermanns U, Sebbel P, Glockshuber R, Haner M, Taschner N, Burkhard P, Aebi U, Muller SA. J Mol Biol. 2002; 323:845. [PubMed: 12417198]
6. Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, Tainer JA. Nature. 1995; 378:32. [PubMed: 7477282]
7. Remaut H, Waksman G. Curr Opin Struct Biol. 2004; 14:161. [PubMed: 15093830]
8. Kang HJ, Baker EN. J Biol Chem. 2009; 284:20729. [PubMed: 19497855]
9. Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN. Science. 2007; 318:1625. [PubMed: 18063798]
10. Scott JR, Zahner D. Mol Microbiol. 2006; 62:320. [PubMed: 16978260]
11. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G. Nat Rev Microbiol. 2006; 4:509. [PubMed: 16778837]
12. Ton-That H, Marraffini LA, Schneewind O. Biochim Biophys Acta, Mol Cell Res. 2004; 1694:269.

13. Ton-That H, Mazmanian SK, Alksne L, Schneewind O. J Biol Chem. 2002; 277:7447. [PubMed: 11714722]
14. Ton-That H, Schneewind O. Mol Microbiol. 2003; 50:1429. [PubMed: 14622427]
15. Ton-That H, Schneewind O. Trends Microbiol. 2004; 12:228. [PubMed: 15120142]
16. Alegre-Cebollada J, Badilla CL, Fernandez JM. J Biol Chem. 2010; 285:11235. [PubMed: 20139067]
17. Hu H, Lu ZY, Parks JM, Burger SK, Yang WT. J Chem Phys. 2008; 128:18.
18. Hu H, Lu ZY, Yang WT. J Chem Theory Comput. 2007; 3:390. [PubMed: 19079734]
19. Hu H, Yang WT. Annu Rev Phys Chem. 2008; 59:573. [PubMed: 18393679]
20. MacKerell AD, et al. J Phys Chem B. 1998; 102:3586.
21. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J Chem Phys. 1983; 79:926.
22. Becke AD. Phys Rev A. 1988; 38:3098. [PubMed: 9900728]
23. Becke AD. J Chem Phys. 1993; 98:5648.
24. Lee CT, Yang WT, Parr RG. Physical Review B. 1988; 37:785.
25. Vosko SH, Wilk L, Nusair M. Can J Phys. 1980; 58:1200.
26. Frisch, MJ., et al. Gaussian 03. C.02. Gaussian, Inc; Wallingford CT: 2004.
27. Hu H, Lu ZY, Yang WT. J Chem Theory Comput. 2007; 3:1004.
28. Lu ZY, Yang WT. J Chem Phys. 2004; 121:89. [PubMed: 15260525]
29. Zhang YK, Lee TS, Yang WT. J Chem Phys. 1999; 110:46.
30. Parks JM, Hu H, Cohen AJ, Yang WT. J Chem Phys. 2008; 129:6.
31. Jónsson, H.; Mills, G.; Jacobsen, KW. Classical and Quantum Dynamics in Condensed Phase Simulations. Berne, BJ.; Ciccotti, G.; Coker, DF., editors. World Scientific; Singapore: 1998. p. 385
32. Burger SK, Yang W. J Chem Phys. 2007; 127:7.
33. Burger SK, Yang WT. J Chem Phys. 2006; 124:13.
34. Hu H, Boone A, Yang WT. J Am Chem Soc. 2008; 130:14493. [PubMed: 18839943]
35. Parks JM, Hu H, Rudolph J, Yang WT. J Phys Chem B. 2009; 113:5217. [PubMed: 19301836]
36. Zeng XC, Hu H, Hu XQ, Yang WT. J Chem Phys. 2009; 130:8.
37. Schaftenaar G, Noordik JH. J Comput-Aided Mol Des. 2000; 14:123. [PubMed: 10721501]
38. Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Proteins: Struct, Funct, Genet. 2003; 50:437. [PubMed: 12557186]
39. Humphreys DD, Friesner RA, Berne BJ. J Phys Chem. 1994; 98:6885.
40. Darden T, York D, Pedersen L. J Chem Phys. 1993; 98:10089.
41. Ryckaert JP, Ciccotti G, Berendsen HJC. J Comput Phys. 1977; 23:327.
42. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. J Chem Phys. 1984; 81:3684.
43. Pickart CM. Annu Rev Biochem. 2001; 70:503. [PubMed: 11395416]
44. Greenberg CS, Birckbichler PJ, Rice RH. FASEB J. 1991; 5:3071. [PubMed: 1683845]
45. Weisel JW. Fibrous Proteins: Coiled-Coils, Collagen and Elastomers. 2005; 70:247.
46. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Science. 2000; 289:2129. [PubMed: 11000116]
47. Heigstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. J Mol Biol. 2003; 334:885. [PubMed: 14643655]
48. Zakeri B, Howarth M. J Am Chem Soc. 2010; 132:4526. [PubMed: 20235501]
49. Lura R, Schirch V. Biochemistry. 1988; 27:7671. [PubMed: 3207697]
50. Hagan RM, Bjornsson R, McMahon SA, Schomburg B, Braithwaite V, Buhl M, Naismith JH, Schwarz-Linek U. Angew Chem Int Ed Engl. 2010; 49:8421. [PubMed: 20878961]

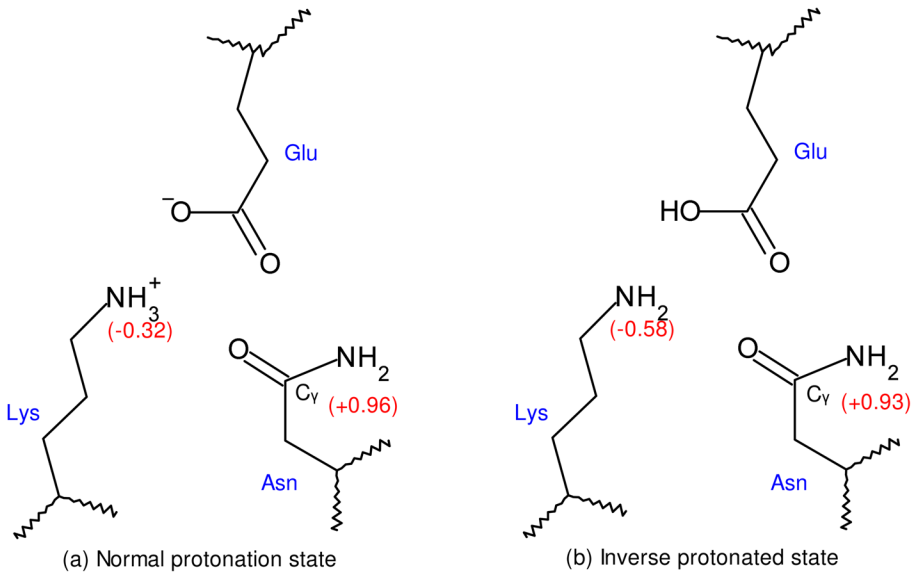


Figure 1.

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

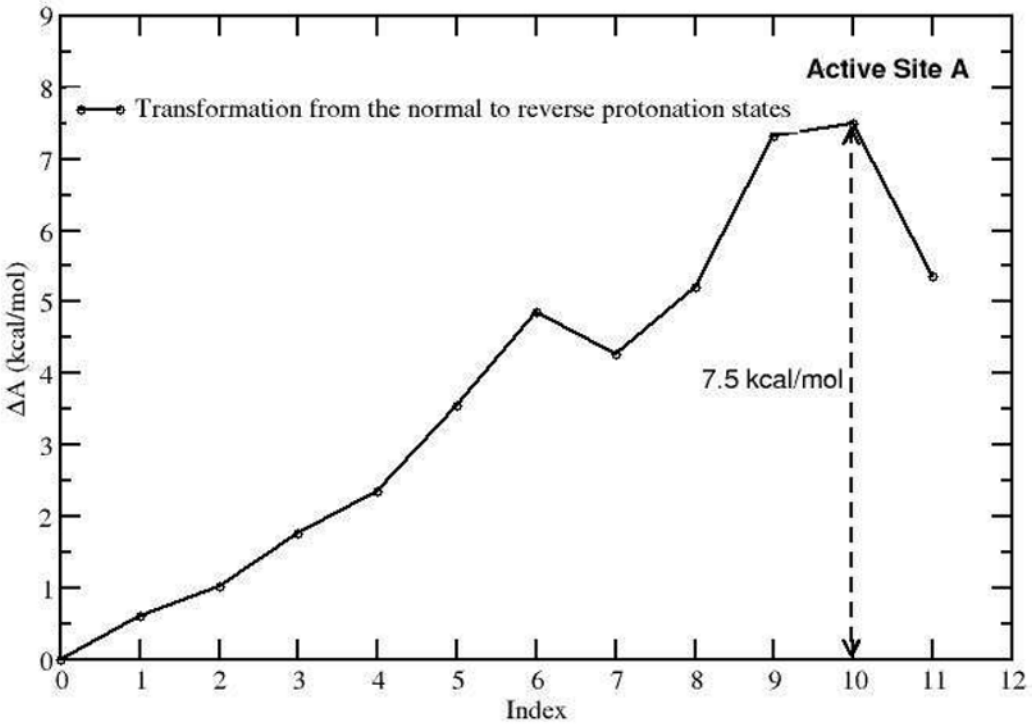
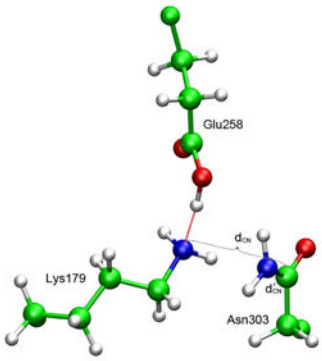
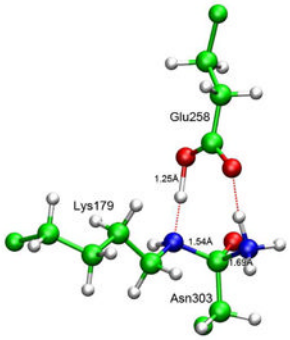


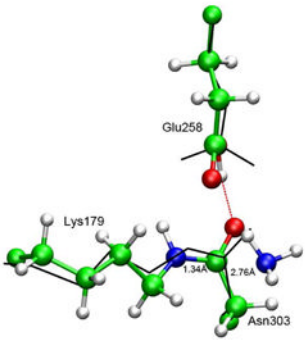
Figure 2.
Potential of mean force of the transformation from normal to *inverse protonated* states for Active Site A.



(a) Reactant



(b) Transition state



(c) Product

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 inter-proton 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.

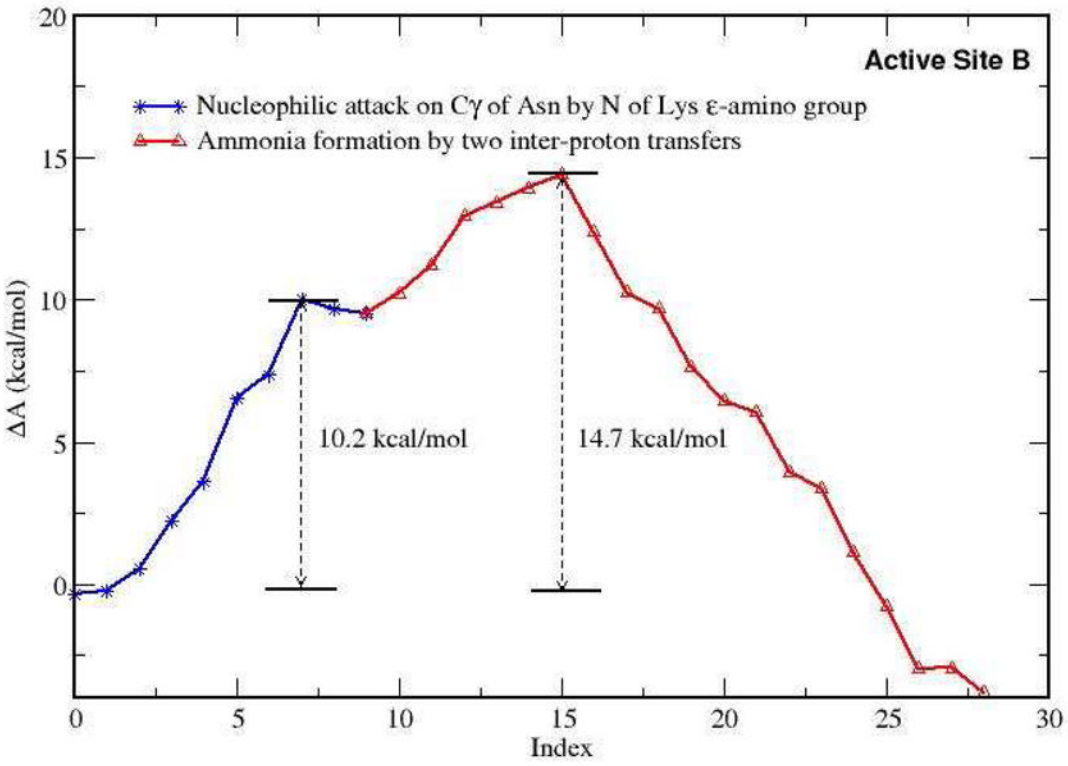


Figure 4.
Potential of mean force of the isopeptide bond formation for Active Site B.

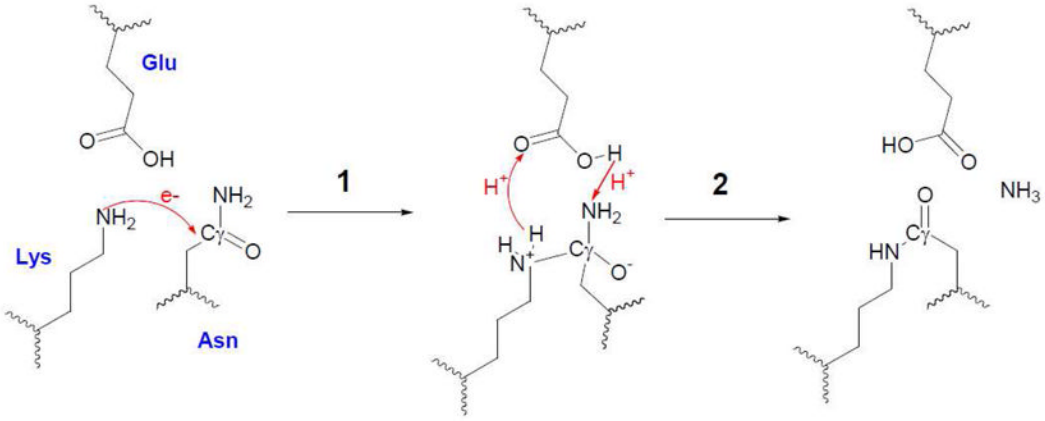
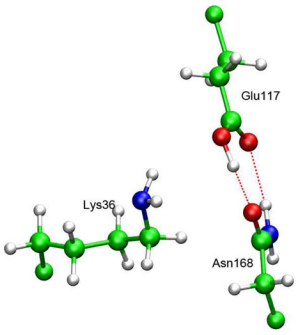
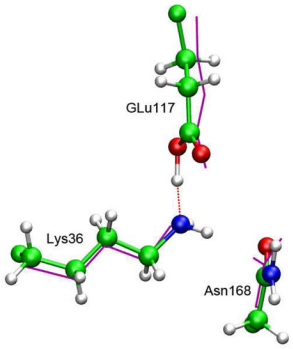


Figure 5.

Putative direct nucleophilic attack reaction mechanism for the isopeptide bond formation in a pilin subunit. The first step is nucleophilic attack on Cγ 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.



(a) Reactant for the wild-type pilus



(b) Reactant for the Pilus_Cut model

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 the purple line.

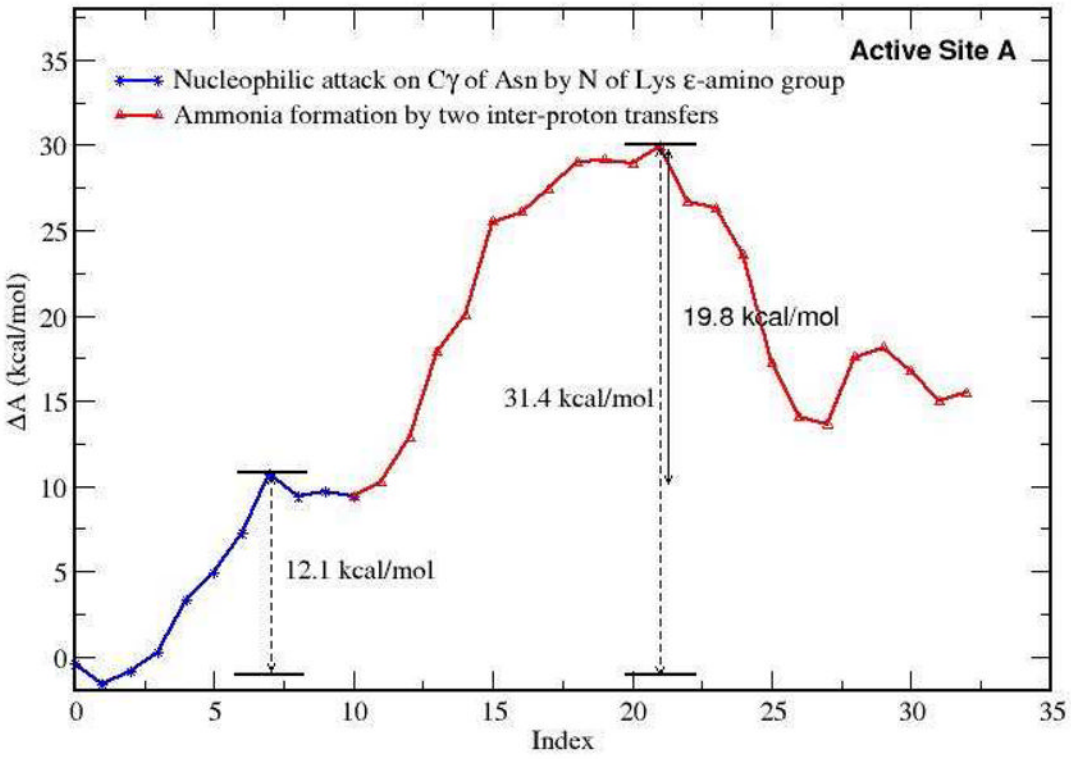


Figure 7.
Potential of mean force of the isopeptide bond formation for Active Site A.

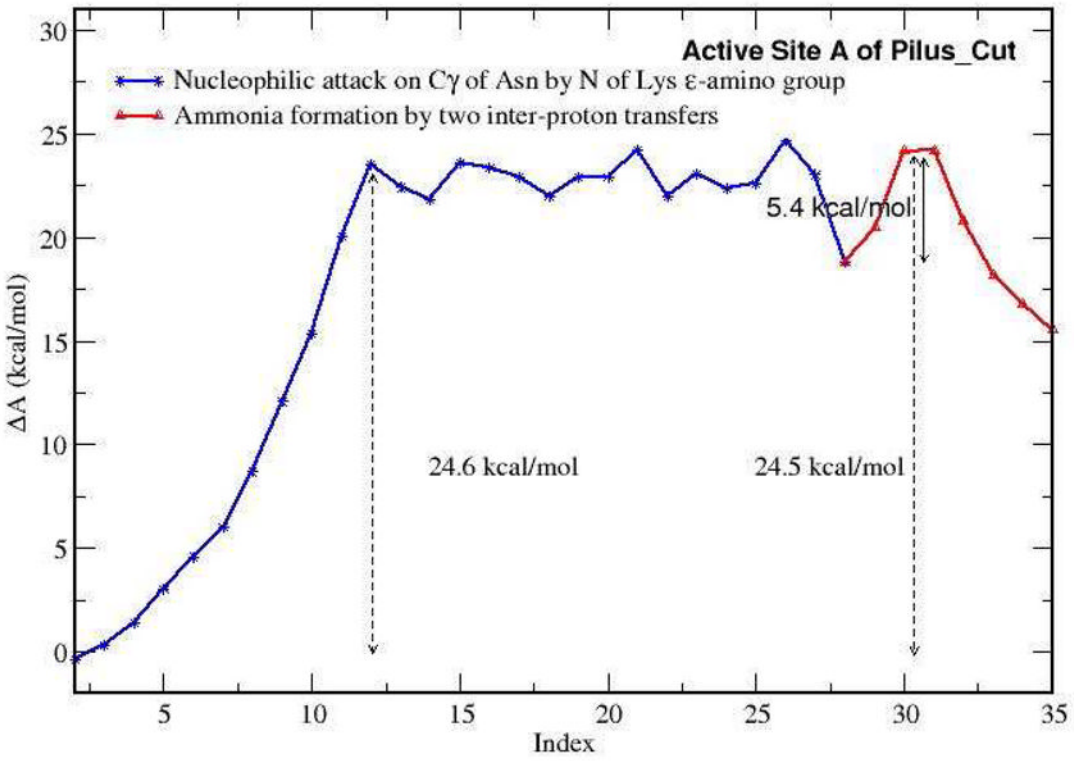


Figure 8.
Potential of mean force of the isopeptide bond formation for Active Site A of Pilus_Cut.