Supporting information for "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

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

1. Supporting figures

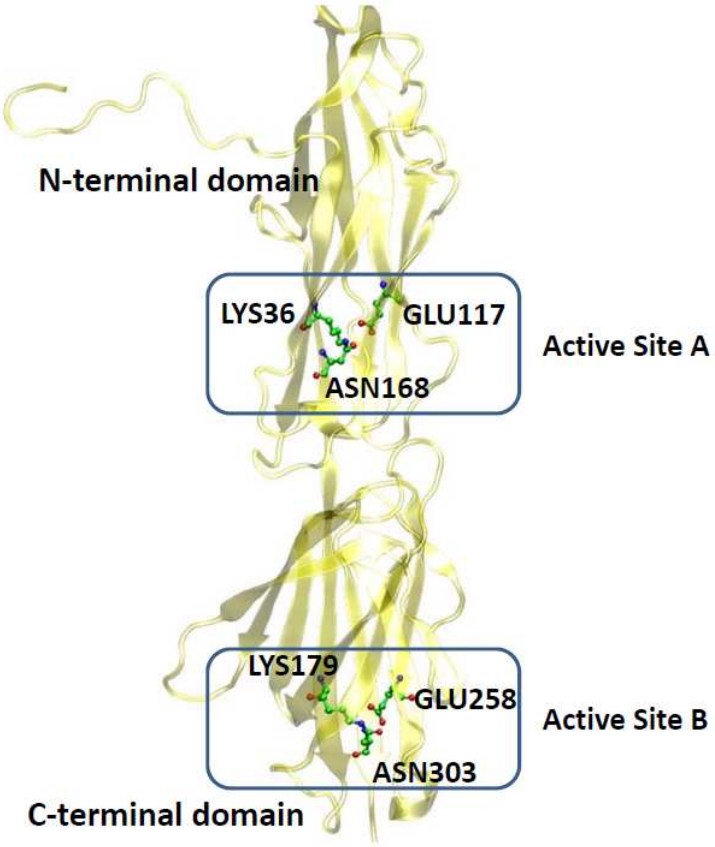
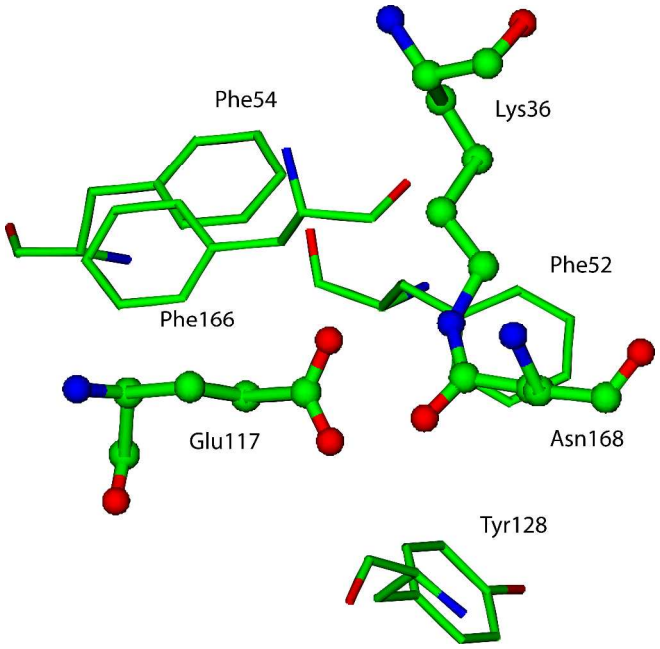
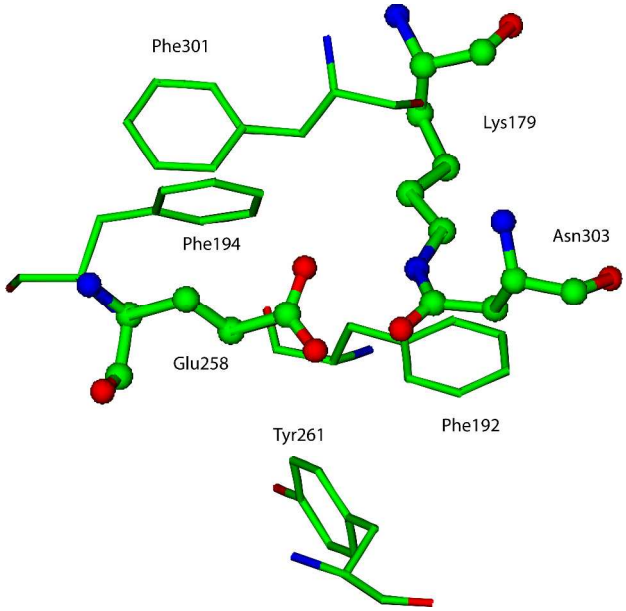


Figure S1. The crystal structure of the pilus (PDB ID: 2IW5). Both active sites include Lys, Asn, and Glu.



(a) The isopeptide bond A.



(b) The isopeptide bond B.

Figure S2. The aromatic amino acids surrounding two isopeptide bonds.

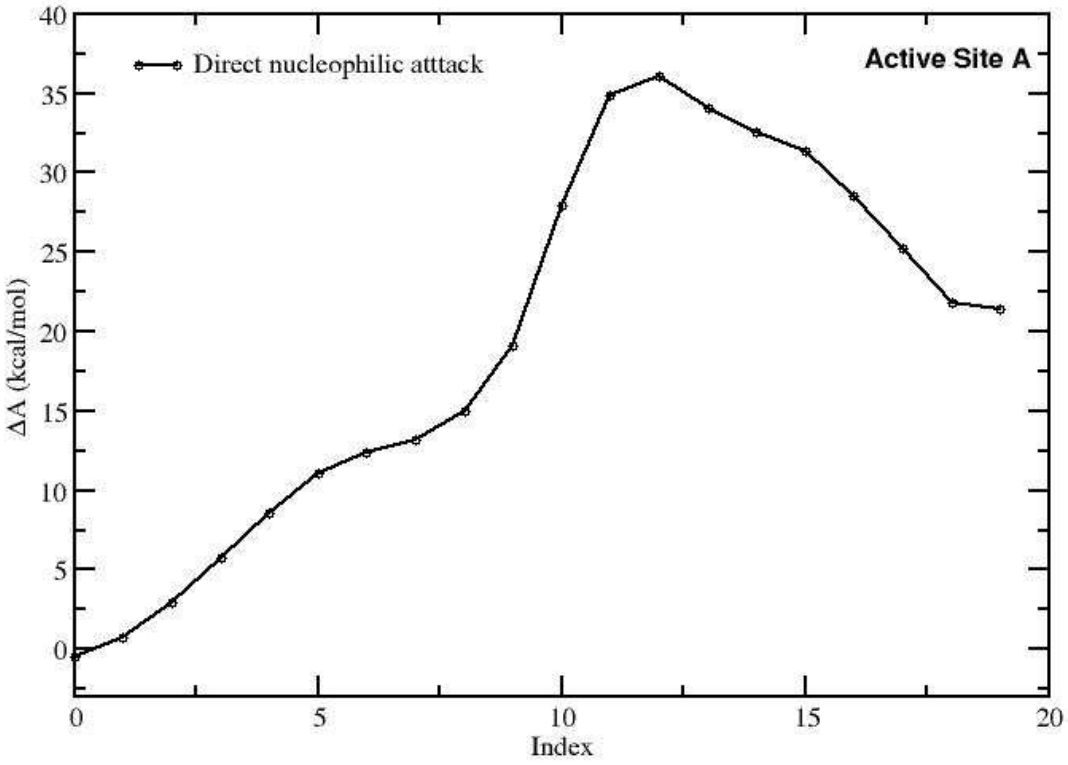


Figure S3. The potential of mean force changes during direct nucleophilic attack on Cγ by nitrogen in the protonated Lys ε-amino group.

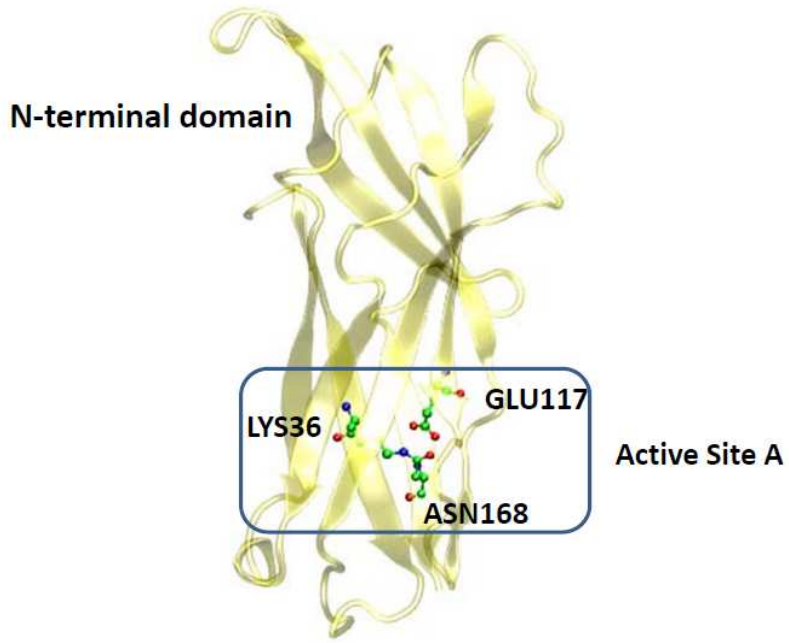


Figure S4. The structure of Pilus_Cut by removing the C-terminal domain and the first 16 residues. The Pilus_Cut model includes the residues from 30 to 171 in the original crystal structure (PDB ID: 2IW5).

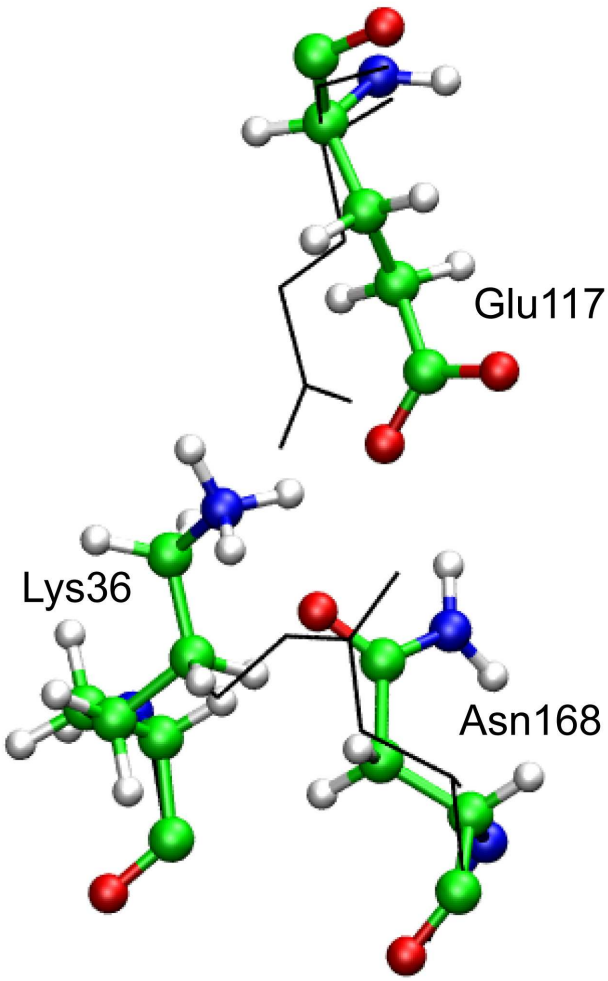


Figure S5. A typical MD snapshot for Active Site A (show in the stick-and-ball style) compared to the crystal structure with the isopeptide bond (shown in the black line). MD simulations were carried out in the norm protonation state without any restraints. Both active sites A and B have shown large fluctuations.

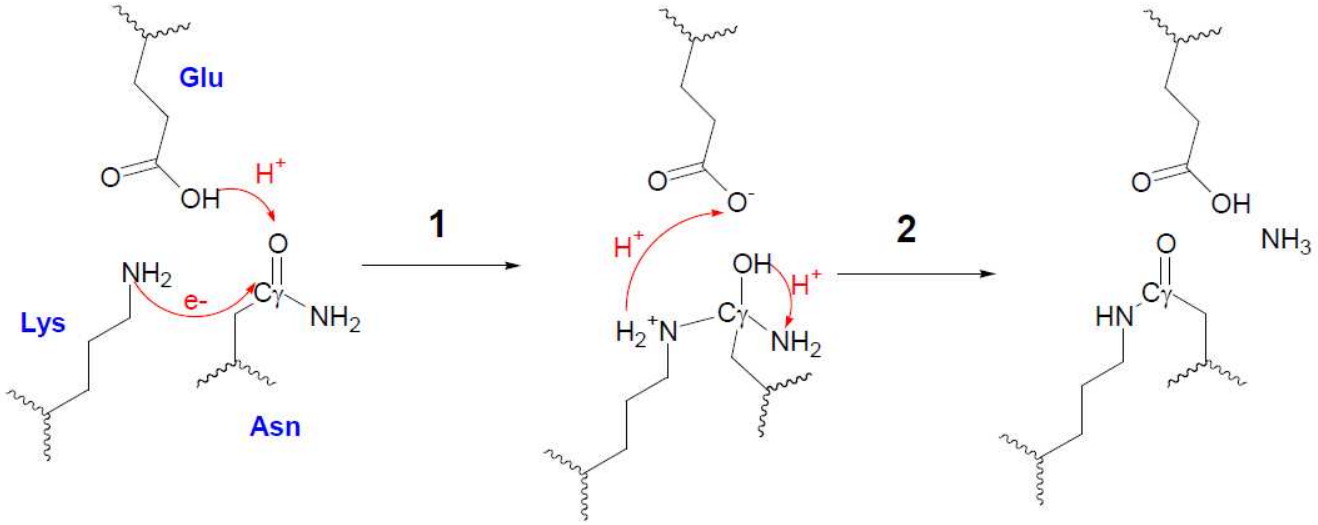


Figure S6. Another reaction mechanism studied in this work. The first step is nucleophilic attack on Cγ in the Asn group by nitrogen in the unprotonated Lys ε-amino group while the proton is transferred from Glu to Asn to generate a stable intermediate. The second step involves two proton transfers to release the ammonia. One of these transfers is intramolecular, which leads to a very high barrier (>50 kcal/mol shown in Figure S8) to form the ammonia in our simulations.

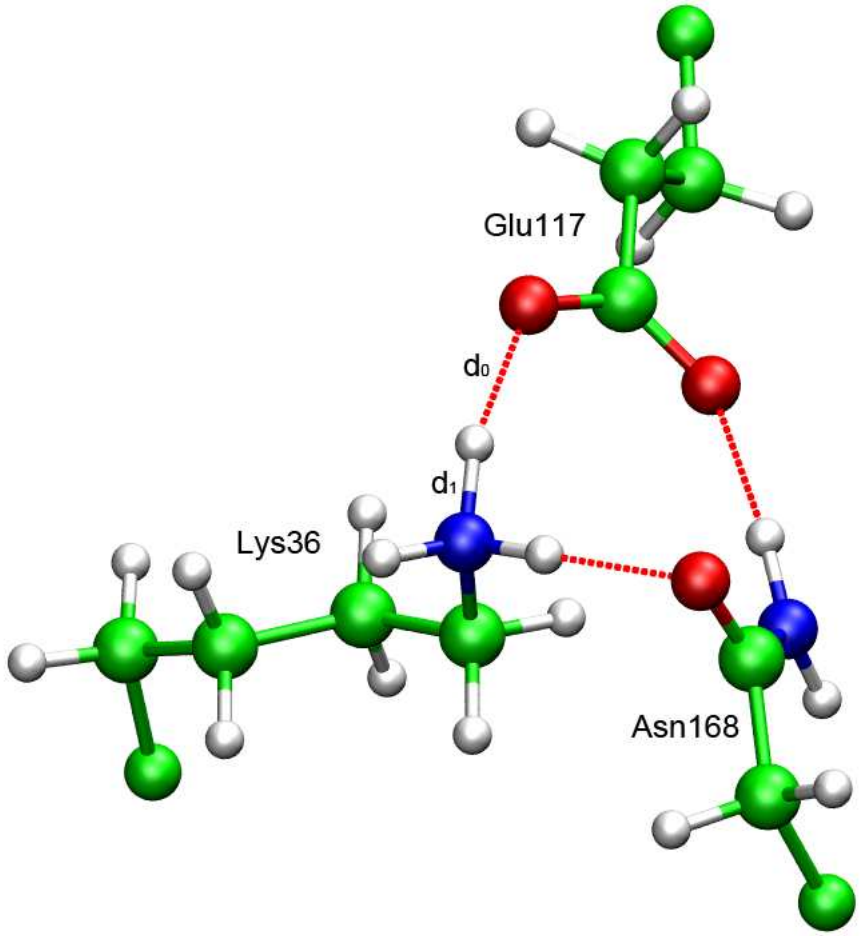


Figure S7. The reactant geometry of Active Site A optimized by the QM/MM-MFEP method. The optimization started from the last snapshot of the 640 ps MD simulation without any restraints. Carbon atoms are in the green color; hydrogen atoms are in the white color; nitrogen atoms are in the blue color; and the oxygen atoms are in the red color. The red dashed lines represent the hydrogen bonds between O and H.

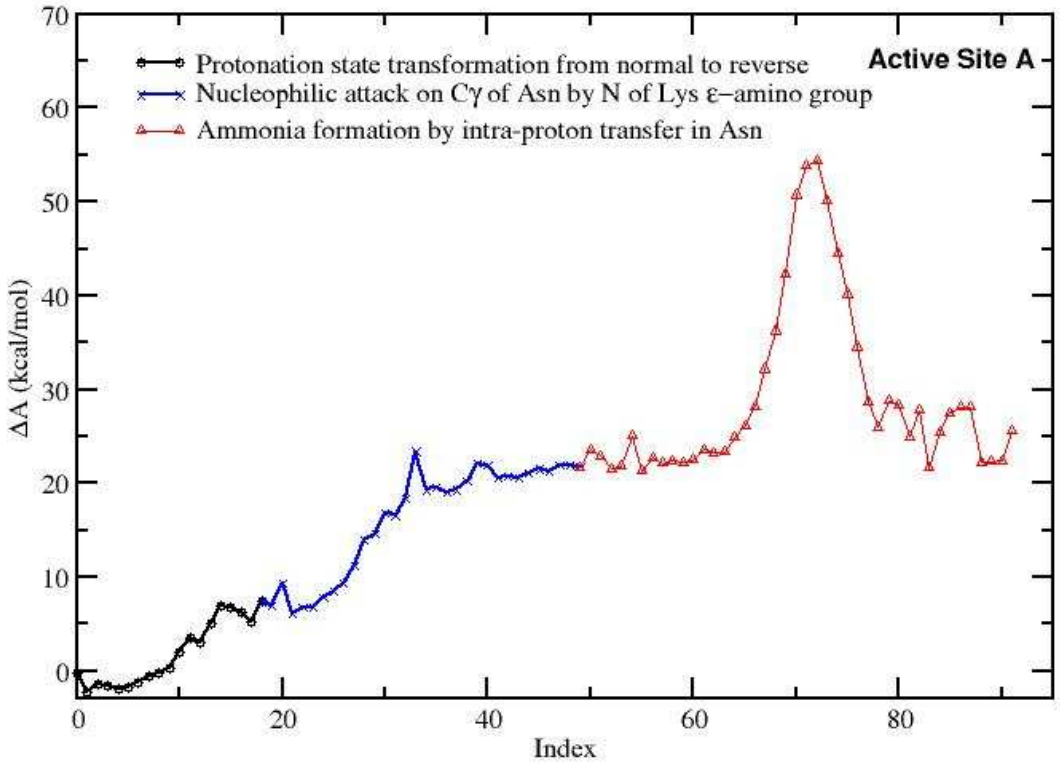


Figure S8. Potential of mean force of the isopeptide bond formation for Active Site A using the scheme in Figure S6.

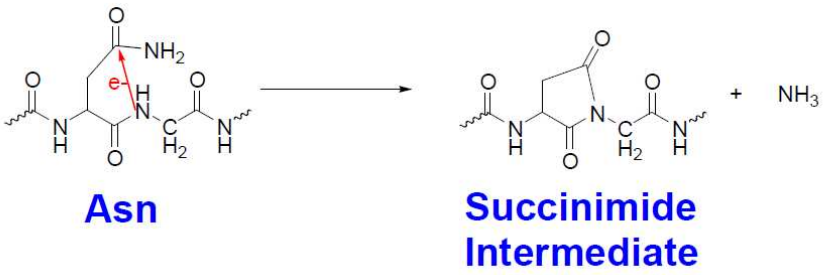


Figure S9. Asparagine can form cyclic succinimide intermediate and release the ammonia. The hydrophobic environment in pili might protect this intermediate from hydrolysis. Then the isopeptide bond between Asn and Lys can be generated later. In this scenario, the ammonia is released first before the isopeptide bond formation. The conversed pilin Glu residues could also be involved in the intermediate succinimide formations or in the isopeptide formations.

2. Another reaction mechanism shown in Figure S6 was studied for Active Site A when the reactant geometry is optimized starting from the last snapshot of the MD simulation without any restrained forces.

When the QM/MM-MFEP geometry optimizations of the reactant state started from the last snapshot of the 640 ps MD simulation without any restraints, Active Site A shown in Figure S7 forms a very strong hydrogen network among Asn^{168}, Lys^{36}, and Glu^{117} while Active Site B only forms one hydrogen bond between Asn^{303} and Glu^{258}. The geometry difference between the active sites A and B is due to the different protein environments between both sites. Active Site B is at the end part of the C-terminal domain and thus might be more flexible than Active Site A. Due to the domain-domain interactions, Active Site A is more rigid than site B. (See Figure S1 for the positions of active sites A and B in the pilus.)

For Active Site A, the driving coordinate for the nucleophilic attack step is the distance (d_{CN}) between Cγ of Asn^{168} and N of the unprotonated Lys^{36} ε-amino group. The driving coordinate for releasing the ammonia is the distance H-N in the intra proton transfer with the distance (d'_{CN}) between Cγ and N of Asn^{168}. The optimized reaction path of the isopeptide bond formation for Active Site A is shown in Figure S8, including the transformation from the normal to *inverse protonated* state.

During the path optimization, the proton in the neutral Glu^{117} group was transferred to the oxygen atom of the carbonyl group of Asn^{168}. This simultaneous proton transfer occurs to form the stable tetrahedral intermediate (shown in the second step of Figure S6). The oxygen atom in Asn^{168} has highly negative charge (around -0.8) after the isopeptide bond is formed. The entire reaction follows exactly the scheme in Figure S6. Although the activation barrier for nucleophilic attack is low (~23 kcal/mol including the first step and relatively ~16 kcal/mol for this step), the last step to release the ammonia through the intra proton transfer from –OH to $–NH_2$ in the tetrahedral intermediate state is forbidden due to the extremely high activation barrier (~54 kcal/mol in Figure S8).

This reaction path calculation indicates that: 1) the activation barrier for the protonation state transformation is low and the nucleophilic attack step is always likely. Thus the reaction path optimizations in our work will always start from the *inverse protonated* state; 2) the intra-residue proton transfer to form the ammonia cannot occur because of the high activation barrier. Several different driving coordinates were used to generate the initial paths and avoid forming the tetrahedral intermediate state. However, when the reaction path optimizations were performed and fully relaxed, the tetrahedral intermediate state is always formed. This indicates that the reaction mechanism in Figure S6 is not plausible.

3. Complete references 20 and 26:

Ref. 20: MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

Ref. 26: Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng,

C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; In *Gaussian 03*; C.02 ed.; Gaussian, Inc.: Wallingford CT, 2004.

4. The coordinates of the atoms in the QM subsystems optimized by QM/MM-MFEP for Figures 3, 6, 7, 8, 10, S5 and S7. The atoms labeled by "F" represent the pseudo-bond atoms.

The reactant coordinates of Active Site B in angstrom for Figure 3:

F	-3.9914987	1.0924147	-29.6204962
C	-3.7659005	-0.3044263	-30.3732908
H	-3.1613171	-0.9265518	-29.7126810
H	-4.7287478	-0.7918902	-30.5249017
C	-3.0365147	-0.1454469	-31.6905240
H	-3.0811569	-1.1067483	-32.2182592
H	-3.5721513	0.5716125	-32.3347841
C	-1.5745358	0.2947793	-31.4964386
H	-1.4637319	0.7102619	-30.4874707
H	-0.9154564	-0.5792953	-31.5487602
C	-1.0876885	1.3481139	-32.4877359
H	-1.0706502	0.9306206	-33.5046553
H	-1.7908027	2.1951773	-32.4876883
N	0.2758917	1.7871735	-32.1298124
H	1.7204301	0.7848083	-32.2814684
H	0.4754876	2.7006956	-32.5391618
H	0.3136212	1.9613808	-31.1237171
F	0.1558561	5.1044349	-35.0733229

C	0.1989412	3.6063904	-35.5571951
H	-0.7100182	3.0786038	-35.2566136
H	0.2829806	3.5357180	-36.6471477
C	1.4326192	2.8824533	-35.0318119
O	2.3847235	3.5280804	-34.6019882
N	1.4540273	1.5409716	-35.1369935
H	0.6429944	0.9881154	-35.3902706
H	2.2154078	1.0476123	-34.6823071
F	5.9932133	-1.0856434	-28.8868891
C	5.0017652	0.0217798	-29.4324415
H	4.2407606	0.2058500	-28.6763945
H	5.5335021	0.9619147	-29.6001707
C	4.2928383	-0.4115674	-30.7046548
H	4.9690339	-0.4507550	-31.5704648
H	3.8766388	-1.4210743	-30.5771439
C	3.1475237	0.5021828	-31.0575374
O	2.7564991	1.4023160	-30.3466306
O	2.5639479	0.2170941	-32.2427948

The transition state coordinates of Active Site B in angstrom for Figure 6:

F	-3.826247	1.246391	-29.524970
C	-3.324432	-0.089873	-30.269125
H	-2.576454	-0.547744	-29.619875
H	-4.165552	-0.772198	-30.381298
C	-2.699337	0.185054	-31.623649

H	-2.569107	-0.776694	-32.137335
H	-3.401198	0.761685	-32.249581
C	-1.351493	0.914514	-31.530343
H	-1.363160	1.606442	-30.682593
H	-0.553424	0.188400	-31.339782
C	-1.022298	1.695080	-32.798072
H	-0.907948	0.996758	-33.634480
H	-1.848077	2.373955	-33.044612
N	0.227031	2.487409	-32.683992
H	1.786638	0.803773	-33.673493
H	0.027271	3.382440	-32.224146
H	1.193702	1.957522	-31.862118
F	-0.117136	5.031903	-34.847880
C	0.037438	3.465766	-35.083134
H	-0.945568	2.990767	-35.138985
H	0.585180	3.307492	-36.015109
C	0.953200	2.850389	-33.988385
O	1.968711	3.568035	-33.733918
N	1.408294	1.444342	-34.530446
H	0.747604	0.898488	-35.092196
H	2.256072	1.631767	-35.091169
F	5.432068	-0.484253	-28.829039
C	4.401825	0.496600	-29.543857
H	3.638390	0.774678	-28.819347
H	4.915242	1.402854	-29.874487

C	3.744967	-0.213401	-30.719513
H	4.490122	-0.528894	-31.457937
H	3.282766	-1.141201	-30.351829
C	2.689652	0.488500	-31.536856
O	2.079349	1.509445	-31.062470
O	2.441557	0.003370	-32.684431

The product coordinates of Active Site B in angstrom for Figure 7:

F	-3.744887	1.306711	-29.553807
C	-3.269345	-0.076360	-30.237987
H	-2.455747	-0.472009	-29.626668
H	-4.102057	-0.779976	-30.222558
C	-2.792336	0.103491	-31.666848
H	-2.648969	-0.891697	-32.105351
H	-3.579337	0.583352	-32.268733
C	-1.494992	0.913613	-31.742365
H	-1.430446	1.549451	-30.850219
H	-0.622472	0.251044	-31.729319
C	-1.372070	1.843073	-32.952002
H	-1.235308	1.270007	-33.876092
H	-2.263713	2.474383	-33.051226
N	-0.193585	2.661911	-32.710584
H	1.597187	0.671695	-34.063052
H	0.156042	2.633846	-31.760247
H	2.469328	2.407392	-31.902553

F	-0.152587	5.318534	-34.814754
C	0.151275	3.739536	-34.908170
H	-0.743488	3.194936	-35.200908
H	0.948905	3.586642	-35.624205
C	0.646243	3.267768	-33.556256
O	1.793045	3.550420	-33.157930
N	1.463658	1.067926	-34.996939
H	1.148968	0.302096	-35.593521
H	2.394806	1.331313	-35.322933
F	5.484486	-0.541095	-28.667010
C	4.506413	0.422502	-29.443202
H	3.731460	0.744773	-28.744713
H	5.033846	1.310489	-29.794067
C	3.820936	-0.267821	-30.610266
H	4.557647	-0.596000	-31.361117
H	3.297531	-1.172048	-30.279449
C	2.835804	0.561232	-31.405341
O	3.001145	1.886349	-31.240156
O	2.014829	0.082544	-32.155239

The reactant coordinates of Active Site A in the inverse protontated state for Figure 8

F	-2.9703302	-3.9173479	12.9523297
C	-1.4206589	-3.8274596	12.5898665
H	-0.9217512	-3.3019764	13.4093922
H	-1.0079763	-4.8386433	12.5329691

C	-1.1510116	-3.1127691	11.2830523
H	-0.0667238	-3.0988388	11.1205400
H	-1.5810834	-3.6849012	10.4461273
C	-1.6492998	-1.6704939	11.2105281
H	-2.7430128	-1.6213124	11.1647348
H	-1.3714429	-1.1331928	12.1308487
C	-1.0737592	-0.9207635	10.0122982
H	-1.2415026	-1.5091163	9.0996602
H	-1.6084473	0.0299025	9.8823439
N	0.3706862	-0.7482315	10.1698778
H	0.5742231	0.0139137	10.8151041
H	-0.0791106	2.4847992	10.0967998
H	0.8344266	-0.5065401	9.2963629
F	-4.5430672	1.5843675	7.4835823
C	-2.9735649	1.6260433	7.5668492
H	-2.5514665	0.6182848	7.5416377
H	-2.5561961	2.1919049	6.7317249
C	-2.3593963	2.2816723	8.7911075
O	-1.1123328	2.3003606	8.8285533
N	-3.1430981	2.8038566	9.7354834
H	-4.1665976	2.6735211	9.7463957
H	-2.6912097	3.1972164	10.5590766
F	1.4946501	4.7022367	15.2258795
C	0.4571313	4.3696292	14.0799113
H	-0.3235082	3.7318958	14.4927742

H -0.0226228 5.2876528 13.7319238

C 1.1024948 3.6516788 12.9081487

H 1.8997004 4.2528812 12.4452030

H 1.5904417 2.7280909 13.2450412

C 0.1240545 3.2959807 11.8188865

O 0.6290010 2.5969042 10.8205702

O -1.0658129 3.6330867 11.8453914

The reactant coordinates of Active Site A in the inverse protontated state of Pilus_Cut for Figure 10:

F -2.804530 -1.648364 -9.268898

C -1.497281 -2.252610 -9.961655

H -0.720226 -2.357748 -9.198045

H -1.723043 -3.242826 -10.365933

C -0.952588 -1.374226 -11.079954

H -0.285097 -1.975563 -11.714303

H -1.769362 -1.048931 -11.742713

C -0.176482 -0.153067 -10.577900

H -0.703639 0.314443 -9.737496

H 0.802540 -0.474506 -10.199968

C 0.008101 0.884450 -11.687105

H 0.158036 0.412292 -12.663789

H -0.859235 1.548428 -11.751982

N 1.212634 1.733331 -11.442058

H 1.236730 2.538410 -12.105466

H 1.277389 2.075800 -10.446130

H	2.086465	1.203552	-11.540966
F	-2.702500	4.510112	-14.269644
C	-1.255610	3.917078	-14.467382
H	-1.275400	2.834620	-14.619673
H	-0.759582	4.365956	-15.334258
C	-0.263685	4.132930	-13.332349
O	0.856358	3.599550	-13.481881
N	-0.600591	4.853190	-12.259342
H	-1.567466	5.179400	-12.089359
H	0.013777	4.819204	-11.428798
F	4.685211	3.774508	-5.708749
C	3.657086	3.677095	-6.944317
H	2.692058	3.361289	-6.550716
H	3.522826	4.654322	-7.421430
C	4.156894	2.648706	-7.932626
H	5.070161	2.994855	-8.434929
H	4.441598	1.750295	-7.373817
C	3.222678	2.155865	-9.023347
O	1.977620	2.351903	-8.913805
O	3.765036	1.526952	-9.992708

The coordinates of Active Site A for Figure S5 from MD simulations:

N	6.783985	2.442777	12.756954
H	7.206592	1.548197	12.879993
C	5.948329	2.477380	11.597048

H	5.702816	3.519412	11.454585
C	6.696035	1.920559	10.373802
H	7.655274	2.473063	10.279337
H	7.013498	0.860887	10.476929
C	5.922678	2.134319	9.071064
H	6.379430	1.323724	8.463909
H	4.855581	1.825411	9.056956
C	5.949434	3.561597	8.520550
H	5.621606	4.384547	9.191075
H	7.039471	3.655066	8.327145
C	5.379429	3.519711	7.101309
H	5.897868	2.758099	6.480432
H	4.310402	3.231886	7.008252
N	5.428832	4.824215	6.404007
H	5.066198	5.589338	7.007898
H	6.382696	5.216093	6.269187
H	4.908573	4.746458	5.506853
C	4.719476	1.666517	11.826229
O	4.765977	0.634146	12.493276
N	10.733114	10.550682	7.113056
H	9.994937	11.178325	6.878175
C	10.613265	9.313290	6.406376
H	11.218516	8.530235	6.838695
C	9.182673	8.749304	6.434158
H	9.058024	8.464329	7.500729

H	8.489551	9.594688	6.236102
C	9.012377	7.539010	5.513785
H	9.073810	7.827852	4.442750
H	9.932976	6.953038	5.722250
C	7.765646	6.695038	5.737080
O	7.744429	5.531266	5.254601
O	6.840184	7.093406	6.493668
C	11.059706	9.484002	4.995119
O	10.509130	10.353328	4.321293
N	0.270381	5.281334	9.072576
H	0.523818	4.346173	9.307635
C	-0.007352	5.451387	7.680151
H	-0.407641	6.419655	7.418187
C	1.238046	5.054724	6.869544
H	1.566167	4.029010	7.142614
H	1.033953	5.158364	5.782380
C	2.343028	6.052288	7.186350
O	3.270250	5.723846	7.924783
N	2.178741	7.302025	6.675700
H	1.555238	7.432285	5.904807
H	3.037021	7.812875	6.724537
C	-1.156315	4.527191	7.466038
O	-1.121112	3.367628	7.874782

The reactant coordinates of Active Site A in the normal protonated state for Figure S7:

F	-0.5140190	2.8893640	-11.7774181
C	0.7848460	1.9663171	-11.3731146
H	0.9172760	1.2880120	-12.2130604
H	1.6416370	2.6364210	-11.2908106
C	0.5675850	1.2126870	-10.0749130
H	1.5231230	0.7374980	-9.8116961
H	0.3438050	1.9124520	-9.2576199
C	-0.5358800	0.1577890	-10.1863699
H	-1.4233360	0.6037070	-10.6457968
H	-0.2158520	-0.6474890	-10.8588514
C	-1.0067641	-0.4500160	-8.8758526
H	-1.3361050	0.3151110	-8.1679592
H	-1.8392490	-1.1232899	-9.0670328
N	0.0534480	-1.2685900	-8.2040138
H	0.8554700	-0.6880140	-7.9008241
H	-0.3278650	-1.7249680	-7.3497982
H	0.3228080	-2.0904281	-8.8227129
F	-4.5534849	-0.5802500	-5.8469691
C	-3.0925810	-1.1787740	-5.4761348
H	-2.3920190	-0.3494560	-5.3907599
H	-3.1488011	-1.7291540	-4.5348320
C	-2.4963970	-2.1419120	-6.4697208
O	-1.3459690	-2.5367110	-6.1852431
N	-3.1757879	-2.4799311	-7.5584488
H	-4.0884719	-2.0591469	-7.7726622

H	-2.7035999	-3.0455430	-8.2798080
F	-1.1545500	-6.2669091	-13.4922380
C	-1.6222240	-5.1984801	-12.3226500
H	-1.8168360	-4.2519131	-12.8228970
H	-2.5345480	-5.5745401	-11.8623600
C	-0.5041440	-5.0410242	-11.2999516
H	-0.3157390	-5.9995942	-10.7903318
H	0.4323220	-4.7892590	-11.8074608
C	-0.7317780	-4.0345101	-10.1736488
O	-1.9003609	-3.8572309	-9.7366829
O	0.3289760	-3.4836321	-9.7238541