Intramolecular Isopeptide Bonds Give Thermodynamic and Proteolytic Stability to the Major Pilin Protein of *Streptococcus pyogenes*.

Received for publication, March 11, 2009, and in revised form, April 28, 2009. Published, JBC Papers in Press, June 4, 2009, DOI 10.1074/jbc.M109.014514

Hae Joo Kang and Edward N. Baker

From the †School of Biological Sciences and the ‡Department of Chemistry, University of Auckland, Auckland 1142, New Zealand

The pili expressed by *Streptococcus pyogenes* and certain other Gram-positive bacterial pathogens are based on a polymeric backbone in which individual pilin subunits are joined end-to-end by covalent isopeptide bonds through the action of sortase enzymes. The crystal structure of the major pilin of *S. pyogenes*, Spy0128, revealed that each domain of the two domain protein contained an intramolecular isopeptide bond cross-link joining a Lys side chain to an Asn side chain. In the present work, mutagenesis was used to create mutant proteins that lacked either one isopeptide bond (E117A, N168A, and E258A mutants) or both isopeptide bonds (E117A/E258A). Both the thermal stability and proteolytic stability of Spy0128 were severely compromised by loss of the isopeptide bonds. Unfolding experiments, monitored by circular dichroism, revealed a transition temperature \( T_m \) of 85 °C for the wild type protein. In contrast, mutants with only one isopeptide bond showed biphasic unfolding, with the domain lacking an isopeptide bond having a \( T_m \) that was \( \sim 30 \) °C lower than the unaltered domain. High resolution crystal structures of the E117A and N168A mutants showed that the loss of an isopeptide bond did not change the overall pilin structure but caused local disturbance of the protein core that was greater for E117A than for N168A. These effects on stability appear also to be important for pilus assembly.

The stability of a globular protein is in general a fine balance between large numbers of weak, noncovalent stabilizing forces (hydrophobic interactions, hydrogen bonds, and ion pairs) and the destabilizing loss of conformational entropy (1). Stability can be greatly enhanced, however, by the presence of a small number of strategically placed covalent bonds, for example those provided by disulfide bonds between Cys residues or by a bound metal ion that bonds to several amino acid side chains. Indeed, some striking examples have been reported in which the introduction of disulfide bonds that link sequentially distant parts of a polypeptide chain can result in large increases in protein stability (2). Our recent discovery of cross-linking isopeptide bonds within the protein subunits of the pili expressed by the Gram-positive organism *Streptococcus pyogenes* (3) has raised questions as to how these bonds are formed and what they contribute to stability. These pili, which are extremely thin (2–5 nm in diameter) but can extend up to 4 μm from the bacterial cell surface (4), are formed as covalently linked polymers through the action of cysteine transpeptidase enzymes called sortases. In this process, which is common to a number of Gram-positive bacterial pathogens, the pilus backbone is formed from a single protein subunit, the so-called major pilin, by the covalent, end-to-end polymerization of major pilin subunits, generating an assembly resembling beads on a string (5–7). Polymerization requires recognition, by the sortase, of an LPXTG-type sequence motif near the pilin C terminus, cleavage after the Thr residue, and transfer to a specific Lys residue in the next pilin subunit. The resulting amide bond, between the terminal COOH of one subunit and the Lys ε-amino group of the next subunit, is referred to as an isopeptide bond.

Surprisingly, the three-dimensional structure of Spy0128, the major pilin from the M1 strain of *S. pyogenes* (group A streptococcus; GAS), revealed the presence of additional isopeptide bonds as internal cross-links within the protein. These bonds, which were confirmed by mass spectrometry, joined lysine and asparagine side chains (3). One such bond was found within each domain of the two domain protein and, in a similar location, between the first and last strands of the domain (Fig. 1). Isopeptide bonds have until now been recognized for their importance in the intramolecular cross-linking of a variety of proteins, such as in ubiquitination (8), transglutamination (9), and sortase-mediated cell wall anchoring of surface proteins (10), as well as pilus polymerization. In this context, the presence of the intramolecular isopeptide bonds in Spy0128 seemed highly unusual. There has been speculation in the past that such internal bonds could exist, but no mechanism has been put forward, and none has previously been proven to exist.

The locations of the two internal Lys-Asn isopeptide bonds in Spy0128 immediately suggested that they are self-generated, each being the result of an intramolecular reaction. Mutagenesis showed that bond formation was in each case dependent on an adjacent Glu residue, Glu117 and Glu258 in the N-terminal domain and Glu258 in the C-terminal domain; mutation of either of these Glu residues to Ala abrogated the formation of the associated bonds (9). From the atomic coordinates and structure factors (codes3GLD and 3GLE) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

† To whom correspondence should be addressed: School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. Tel.: 64-9-3737599; Fax: 64-9-3737619; E-mail: ted.baker@auckland.ac.nz.

‡ The abbreviations used are: GAS, group A streptococcus; ESI-TOF, electrospray ionization-time of flight; WT, wild type; MOPS, 4-morpholinopropanesulfonic acid; N domain, N-terminal domain; C domain, C-terminal domain.
Contributions of Isopeptide Bonds to Bacterial Pilin Stability

FIGURE 1. Overall structure and topology of Spy0128 with internal isopeptide bonds. A, the structure of Spy0128 is shown in ribbon representation with the N terminus (N) and C terminus (C) marked. The isopeptide bond in each of the N and C domains is shown as spheres with residue labels next to them. The residue forming the intersubunit bond, Lys161, is shown in stick mode. The sortase recognition motif, EVPTG, is shown as a dotted line at the C terminus of the structure. B, the topology diagram of Spy0128 is color-coded in rainbow from blue (N terminus) to red (C terminus). The position of Lys161 is indicated by the red arrowhead, and the positions of the isopeptide bonds are indicated by black bars. C, the N domain isopeptide bond of Spy0128 between Lys36 and Asn168 and the catalytic residue Glu117 are shown in ball-and-stick mode.

isopeptide bond (3). A close parallel exists in the self-generated Lys-Asn isopeptide bonds that form during maturation of the protein coat of the bacteriophage HK97, where a nearby Glu residue is essential for the reaction, and the capsid subunits become covalently linked to form interlocked circular rings referred to as chain mail (11, 12).

Further investigations have shown that similar isopeptide bonds can be found as internal cross-links in other proteins. Examination of previously determined structures of a minor pilin GB552 from Streptococcus agalactiae (13) and the CnaA and CnaB domains of a collagen-binding adhesin from Staphylococcus aureus (14–16) revealed constellations of Lys-Asn isopeptide bonds similar to those that form the internal cross-links in Spy0128, and examination of the electron density confirmed their probable presence in those cases where the x-ray data were available (3). Sequence comparisons showed that similar domains, with corresponding Lys-Asn-Glu/Asp residues similar to those that form the internal cross-links in Spy0128, are present in many cell surface proteins of Gram-positive bacteria. Recently the Bacillus cereus major pilin BcpA was shown, by mass spectral analyses, to contain internal isopeptide bonds similar to those in Spy0128 (7), and sequence comparisons point to similar bonds in the major pilins of other species. These data suggest that isopeptide bond cross-links could be important features in many surface proteins involved in adhesive functions, where stability against physical and chemical stresses is important.

Here we describe the preparation of mutants of Spy0128 that lack one or both of the internal isopeptide bonds, using mass spectrometry to confirm their absence. We show by x-ray crystallography that the overall structure of Spy0128 is not significantly affected by the loss of an isopeptide bond. We also show, however, that the proteolytic and thermal stability of Spy0128 is severely compromised when the internal isopeptide bonds are removed and thereby establish their important stabilizing role in proteins of this type.

EXPERIMENTAL PROCEDURES

Cloning of Spy0128—The DNA sequence encoding amino acids 18–311 of Spy0128 was amplified from GAS serotype M1 genomic DNA, using the primer pairs listed in Table 1. The amplified DNA product was double-digested with restriction enzymes EcoRI and BamHI and ligated into the vector pGEX3C. The ligation reaction was transformed into Escherichia coli DH5α cells, and positive colonies were screened by colony PCR. DNA sequencing was carried out to confirm the sequence. The resulting plasmid pGEX3C-Spy0128 contained the desired Spy0128 sequence, a linker with sequence 5’-CAGGGGATCCATCTTGAGACG-3’. The amplified DNA product was double-digested with restriction enzymes EcoRI and BamHI and ligated into the vector pGEX3C. The ligation reaction was transformed into Escherichia coli DH5α cells, and positive colonies were screened by colony PCR. DNA sequencing was carried out to confirm the sequence. The resulting plasmid pGEX3C-Spy0128 contained the desired Spy0128 sequence, a linker with sequence GPGS, a picornavirus 3C protease cleavage site, and a glutathione S-transferase tag.

Site-directed Mutagenesis—Mutations were introduced to Spy0128 by the PCR-based site-directed mutagenesis of double-stranded DNA. Forward and reverse primer pairs for each mutation were designed, as listed in Table 1. Two alternative wild type constructs were used as templates for PCR amplification: pGEX3C-Spy0128, prepared as described above, and pGEX3C-Spy0128 prepared as described earlier (3). After amplification with Pfu Turbo DNA polymerase (Stratagene), each reaction was incubated with DpnI restriction enzyme at 37 °C to remove template DNA and then transformed into E. coli DH5α cells. The sequences of the resulting constructs were confirmed by DNA sequencing.

Expression and Purification of Spy0128 and Mutants—The plasmids for pGEX3C-Spy0128, Spy0128 and their mutants were transformed into E. coli BL21 (DE3) pRP cells for expression. Protein overexpression was carried out in ZYP-5052 autoinduction medium (17). The cells were grown initially at 37 °C for 4 h followed by 20 h at 28 °C. Harvested cells were lysed in a buffer (buffer A) containing 25 mM Tris–HCl, pH 8.0, 50 mM NaCl using sonication or a cell disruptor (Constant Systems). Cell debris was removed by centrifugation, and the final supernatant was filtered through a 0.2-μm filter prior to protein purification.

### Table 1

| Primer | Sequence 5’ → 3’ |
|--------|------------------|
| pGEX3C-Spy0128E117A | CCGGTATTTTACGAGAAAAATTCTAGATCAG |
| Spy0128E117A | GAATTCTGACAGGCCATGGAATTTGCAAG |
| Spy0128N168A | CAAATGCAGAGAGTTATCAAGA CTAGTCTGCAGTGACAACATAATCCAC |
| Spy0128N168A | CAAATGCAGAGAGTTATCAAG |
| Spy0128E258A | CAAATGCAGAGAGTTATCAAG |

Mutated codons are underlined.
Spy0128<sub>18–311</sub> and Spy0128<sub>18–308</sub> were purified as described previously (3). Briefly, the proteins were first purified using a glutathione-Sepharose column (GE Healthcare) followed by cleavage of the glutathione S-transferase tag on the column using recombinant picornavirus 3C protease. The untagged protein was further purified by anion exchange chromatography using a HiTrap QFF column with a NaCl gradient (0–1 M) in buffer A, followed by size exclusion chromatography using a Superdex 200 HR10/30 gel filtration column (GE Healthcare) in buffer A. For the mutant proteins, purification was carried out at 4 °C, and the duration for the glutathione S-transferase tag cleavage was reduced to 4 h instead of overnight, because some mutants were susceptible to degradation. Gel filtration fractions containing pure Spy0128<sub>18–311</sub> or mutants were pooled and concentrated, and protein concentrations were spectrophotometrically determined using the extinction coefficient of 26,360 M<sup>−1</sup> cm<sup>−1</sup> at 280 nm.

**Mass Spectrometry Analyses**—Accurate molecular masses of proteins were determined by infusion ESI-TOF mass spectrometry undertaken on a Q-STAR XL hybrid tandem mass spectrometry system (Applied Biosystems) in 50% acetonitrile with 0.1% formic acid. The raw mass spectrometry data were deconvoluted using the Bayesian protein reconstruction tool in BioAnalyst software (Applied Biosystems).

**CD Spectroscopy**—CD spectroscopy experiments were conducted on a JASCO J-815 CD spectrophotometer equipped with a Peltier temperature control (JASCO Inc.). The proteins were used at a concentration of 0.1 mg/ml ( ~ 3 × 10<sup>−6</sup> M), in 10 mM potassium phosphate buffer, pH 8.0. For CD data collection, 1 ml of each protein was used in a 1-cm path length quartz cuvette. Wavelength scans between 200 and 250 nm were collected at 0.5 °C intervals. Reversibility of thermal denaturation was checked by cooling and reheating the same samples. The raw data in milldeg units were corrected for background absorbance and converted to molar ellipticity [θ]. Thermal denaturation curves were determined by plotting changes in CD molar ellipticity as a function of temperature. The experimental curves were fit to the Boltzmann equation describing a two-state folding transition, and the melting temperature <i>T<sub>m</sub></i> was obtained from the inflection point of the curve.

**Proteolysis Assay**—Spy0128<sub>18–311</sub> and mutant proteins were diluted in 25 mM NH<sub>4</sub>HCO<sub>3</sub> to 0.3 mg/ml and mixed with 0.3% (w/w) trypsin. The reaction was carried out at 37 °C. At different time points between 1 min and 16 h from incubation, 15 μl of each reaction was drawn out and mixed with boiling SDS sample buffer. The samples were analyzed by running on a 15% SDS-PAGE gel followed by Coomassie Blue staining.

**Crystallization and X-ray Data Collection**—Crystallization trials were performed at 18 °C by the vapor diffusion sitting drop method. In each case equal volumes of protein (concentration, 100 mg/ml in 25 mM Tris-HCl, pH 8.0, 50 mM NaCl) and precipitant were mixed using a Cartesian nanoliter dispensing robot (Genome Solutions). The best crystals were obtained for mutants of the Spy0128<sub>18–308</sub> construct. The E117A mutant produced crystals with a precipitant comprising 0.2 M potassium acetate and 20% polyethylene glycol 3350, whereas for the N168A mutant, the best crystals were obtained using 0.2 M MOPS/KOH, pH 7.5, and 21% polyethylene glycol 6000. Prior to data collection, the crystals were soaked for 5 s in cryoprotectant solutions comprising mother liquor solution supplemented with 0–20% ethylene glycol (E117A) or polyethylene glycol 400 (N168A) and were then immediately cryo-cooled in liquid nitrogen. Diffraction data were collected to 2.0 Å resolution at 100 K using the MicroMax-007HF rotating anode generator (Rigaku) equipped with a Mar345 image plate detector. The data sets were processed and scaled with the HKL2000 package (18).

**Data Collection and Structure Determination**—The crystals of both the E117A and N168A mutant proteins belong to space group P2<sub>1</sub>, with three molecules/asymmetric unit and unit cell dimensions very similar to those of the WT crystals. The E117A crystals were isomorphous with WT Spy0128<sub>18–308</sub> but N168A crystals showed slightly altered unit cell dimensions (Table 2). The WT Spy0128<sub>18–308</sub> structure (Protein Data Bank code 3B2M) was used as a phasing model for structure determination after omitting all solvent molecules. The isomorphism of the E117A mutant meant that this WT model could be placed into the E117A unit cell, and its position and orientation could be optimized by rigid body refinement with REFMAC (19). For the less isomorphous N168A crystals, molecular replacement with PHASER (20) was necessary, however, using the WT structure as search model. In both cases, the <i>R<sub>free</sub></i> set of reflections used in the wild type structure refinement was transferred to the mutant reflection file before refinement commenced to avoid bias. Refinement was with REFMAC with TLS restraints (21), and manual model building was carried out using COOT (22). The E117A structure was refined at 2.03 Å resolution to final values of <i>r</i> = 0.212 and <i>R<sub>free</sub></i> = 0.260, and N168A was refined at 2.20 Å resolution to <i>r</i> = 0.205 and <i>R<sub>free</sub></i> = 0.278. For both structures, over 90% of the residues are in the most favored regions of the Ramachandran plot, as defined by PROCHECK (23) with no outliers. The data collection and refinement statistics are given in Table 2. The atomic coordinate and structure factor files were deposited with the RCSB Protein Data Bank with accession codes 3GLD (E117A) and 3GLE (N168A).

**RESULTS**

**Generation of Isopeptide Bond Mutants of Spy0128 and Characterization by ESI-TOF Mass Spectrometry**—The formation of the two internal isopeptide bonds in Spy0128 is dependent in each case on an associated acidic residue, Glu<sup>117</sup> for the N-terminal domain isopeptide bond and Glu<sup>258</sup> for the C-terminal domain isopeptide bond. These residues were mutated to Ala to generate the single-mutation proteins E117A and E258A, and the double-mutation protein E117A/E258A. In addition, Asn<sup>168</sup>, which forms the isopeptide bond with Lys<sup>36</sup> in the N-terminal domain, was also changed to alanine, creating the N168A mutant, to compare with the glutamate substitution mutants. The appropriate mutations were made in a construct encoding residues 18–311 (Spy0128<sub>18–311</sub>), which ends with Thr<sup>311</sup>, known to be the C-terminal residue after sortase cleavage. For crystallography, however, the slightly shorter construct
Contributions of Isopeptide Bonds to Bacterial Pilin Stability

Spy012818–308 was used as this gave much better crystals and allowed direct comparison with the WT structure (3).

Both wild type (WT) Spy0128 and the mutant proteins were expressed and purified, and the presence or absence of isopeptide bonds was analyzed by ESI-TOF mass spectrometry (Table 3). The molecular mass of WT Spy0128 was 34.8 Da less than that calculated from the amino acid sequence (Mcalculated), indicating the loss of two units of NH3 through formation of two isopeptide bonds. For the single-Glu mutants E117A and E258A, loss of only one NH3 was observed in each case, indicating formation of one isopeptide bond. In contrast, no mass decrease was detected for the double mutant E117A/E258A, indicating no isopeptide bonds. These results confirm the essential role of the Glu residues in the isopeptide bond formation. Mutation of Asn168 to Ala in N168A also effectively removed the isopeptide bond, as shown by the loss of only one NH3 from the overall protein mass.

CD Spectra of Spy0128 and Its Isopeptide Bond Mutants—The secondary structures of Spy0128 and its mutants were estimated by CD spectroscopy. The crystal structure of Spy0128 shows that it is an all-β protein with no α-helices (3). The CD spectrum at pH 8.0 (Fig. 2) displayed a distinctive minimum at 215 nm and positive peaks at ~200 nm (not shown) and ~227 nm. These features differ slightly from those seen for classic anti-parallel β-sheet proteins, which show positive bands at ~195 nm and a minimum at ~218 nm. The mutant proteins all displayed similar CD spectra to WT Spy0128, with only minor differences, mainly around the 215-nm region. The CD spectra of E117A and E258A were virtually identical, and both showed deeper 215-nm minima relative to WT, whereas N168A showed a lesser 215-nm minimum than WT. For the double mutant E117A/E258A, the overall spectra were shifted downward. The spectra are consistent with similar folding in all cases.

Thermal Stabilities of Spy0128 and Its Isopeptide Bond Mutants—The thermal stabilities of Spy0128 and its mutant proteins were probed by measuring the temperature dependence of their CD spectra in the far-UV region over a temperature range of 20–95 °C. Unfolding of the proteins was monitored by the loss of ellipticity at 222 nm. This wavelength was chosen because the spectra were often noisy around the 215-nm minimum. The reversibility of unfolding was tested by reheating the same samples and cooling back to the starting temperature, 20 °C. After heating to 95 °C, both WT and N168A proteins could be refolded with no apparent sign of aggregation (Fig. 2, B and D). In contrast, E117A, E258A, and

### Table 2

Data collection and refinement statistics

| Protein          | Spy012818–308 E117A | Spy012818–308 N168A |
|------------------|---------------------|--------------------|
| Space group      | P21                 | P21                |
| a = 67.31, b = 52.15, c = 127.32, β = 98.83° | a = 64.71, b = 50.87, c = 123.20, β = 104.27° |
| Resolution range (Å)<sup>a</sup> | 50–2.03 (2.10–2.03) | 50–2.20 (2.28–2.20) |
| Wavelength (Å)   | 1.54178             | 1.54178            |
| Total reflections | 2079391             | 2274826            |
| No. unique reflections<sup>a</sup> | 54403 (4811)        | 40313 (4012)       |
| Redundancy<sup>a</sup> | 5.9 (4.7)           | 10.4 (10.0)        |
| Completeness<sup>a</sup> | 95.8 (84.9)         | 99.8 (100.0)       |
| Mean I/σ(I)<sup>a</sup> | 32.6 (5.8)          | 38.7 (6.6)         |
| R<sub>sym</sub> (%)<sup>a</sup> | 4.9 (27.8)          | 6.0 (39.5)         |

### Table 3

ESI-TOF mass spectral analyses of wild type and mutant proteins of Spy0128

| Protein          | Expected M<sub>average</sub><sup>a</sup> Da | Observed by ESI-TOF M<sub>average</sub><sup>a</sup> Da | Δ (M<sub>expected</sub> − M<sub>observed</sub>)<sup>b</sup> | NH3 units lost |
|------------------|--------------------------------------------|-----------------------------------------------|---------------------------------------------|---------------|
| Spy012818–311    | 32,756.4                                  | 32,721.6                                      | 34.8                                        | 2             |
| Spy012818–311 E117A | 32,698.3                                | 32,682.0                                      | 16.3                                        | 1             |
| Spy012818–311 N168A | 32,713.3                                | 32,696.5                                      | 16.8                                        | 1             |
| Spy012818–311 E258A | 32,698.3                                | 32,681.5                                      | 16.8                                        | 1             |
| Spy012818–311 E117A/E258A | 32,640.3                              | 32,640.5                                      | 0.2                                         | 0             |

<sup>a</sup> Average molecular mass.

<sup>b</sup> Difference between expected molecular mass and observed molecular mass.
E117A/E258A all precipitated with increasing temperature and showed significant loss of CD signal after heating and cooling, indicating irreversible thermal denaturation of these mutants. In particular, E117A/E258A showed a nearly complete loss of the characteristic CD maximum and minimum. These results indicate that the reversibility of Spy0128 thermal unfolding is significantly perturbed by the Glu to Ala mutations. N168A also lacks the N domain isopeptide bond, like E117A, but it does refold after thermal unfolding. This indicates that the presence of the Lys-Asn isopeptide bonds in Spy0128 is not solely responsible for the reversibility of its thermal unfolding.

All of the proteins showed a gradual decrease in their 222-nm CD signal as the temperature was increased. This was more pronounced in the E117A, E258A, and E117A/E258A mutants, whereas the wild type and N168A showed much slower descent (Fig. 3). The WT protein followed a typical two-state unfolding mechanism, with a sharply defined transition temperature ($T_m$) of 85.1 °C and no evidence of a stable intermediate (Fig. 3). In contrast, all of the single mutants exhibited two transitions, the
first between ~50 and 60 °C and the second between ~80 and 85 °C, similar to the single transition seen for WT (Table 4). The transitions that occurred in the mutant proteins were less sharp than that of WT, however, indicating less cooperative unfolding reactions. The CD signals of the glutamate mutants E117A and E258A decreased more rapidly than that of the N168A from the starting temperature. This indicates that the unfolding was well under way at the low temperatures for the Glu mutants, whereas it was less so for N168A. For the double mutant E117A/E258A, the CD signal decreased markedly beyond 50 °C and then stopped at 60–70 °C, and no post-transition base line could be established.

Proteolytic Stability of Spy0128 and Its Isopeptide Bond Mutants—It has been known for decades that GAS Lancefield T antigens are resistant to trypsin digestion (24). Because Spy0128 has been recently recognized as a Lancefield T1 antigen (4), tests were carried out to determine whether the intramolecular isopeptide bonds play a significant role in the resistance to proteolysis of this protein. To this end, Spy012818–311 and its mutant proteins were subject to trypsin digestion at 37 °C, with the progress of the digestion being followed by SDS-PAGE analysis of samples taken at various time points (Fig. 4).

The WT protein remained essentially intact after 4 h digestion and was still largely intact after 16 h. In contrast, the mutant proteins showed significantly increased susceptibility to proteolysis. No full-length E117A/E258A could be detected after 5 min of digestion, and the single mutants showed varying degrees of proteolytic stability, with the most stable being N168A; most of the N168A protein remained intact after 4 h of digestion, although it was completely digested after 16 h. Like N168A, E117A lacks the N domain isopeptide bond, yet it was digested much more readily, with almost all full-length protein being digested after 2 h. Similar results were obtained with E258A, which lacks the C-terminal domain isopeptide bond. Because the full-length E258A band disappeared during digestion, there was a concomitant intensification of a ~16-kDa band that was subsequently digested away between 4 and 16 h.

Overall Structures of Spy0128 Mutants—The crystal structures of the E117A and N168A mutants, determined at 2.0 Å and 2.2 Å resolution, respectively, showed that both mutants had overall structures that were essentially identical to that of the WT protein. Pairwise structural superpositions of the mutant and WT proteins, each having three independent copies in the crystal asymmetric unit, gave root mean square differences of 0.3–1.2 Å in all Ca atom positions, with the only significant differences being local ones at the mutation sites. These are described below.

**TABLE 4**

| Protein                  | $T_m$ °C |
|--------------------------|---------|
| Spy012818–311            | 85.1    |
| Spy012818–311 E117A      | 50.2/84.2 |
| Spy012818–311 N168A      | 60.3/83.6 |
| Spy012818–311 E258A      | 57.5/79.4 |
| Spy012818–311 E117A/E258A| nd*     |

* nd, not determined.

**FIGURE 4. Proteolytic stability of Spy0128 and mutants.** Trypsin digestion products at various time points were separated on 15% SDS-acrylamide gels. The time points are shown on the left-hand side, and the molecular weight markers are shown on the right-hand side of the gels. Gel lanes from different gels were extended or compressed vertically according to the molecular weight markers and arranged side by side, separated by white space.

**Spy0128 E117A Structure**—The Glu → Ala mutation site was immediately evident in electron density difference maps as a large negative electron density feature, signifying the loss of the Glu117 carboxymethyl moiety (Fig. 5A). Adjacent to this, the two residues that form the N-terminal domain isopeptide bond in WT Spy0128, Lys36 and Asn168, have discrete, unconnected density features, signifying that no isopeptide bond is present. These were the same in all three independent monomers. In contrast, in the C-terminal domain, clear continuous electron density extended over the side chains of Lys177 and Asn303 in each molecule, indicating that the isopeptide bond in the C domain, with nonmutated Glu258, remained intact as in the wild type structure (Fig. 5B). This agrees with the mass spectrometry result showing the presence of only one isopeptide bond in Spy0128 E117A.

In the absence of the catalytic Glu, and isopeptide bond formation, the side chains of Lys36 and Asn168 are free to take alternative orientations. These were found to be somewhat different in each of the three independent molecules, probably reflecting the lack of appropriately positioned hydrogen bonding partners in what is a predominantly hydrophobic environment. The side chain of Asn168 is in essentially the same place as...
Contributions of Isopeptide Bonds to Bacterial Pilin Stability

in WT, albeit with slight rotations of the carboxyamide group, whereas its Oδ1 atom remains hydrogen-bonded to the main chain NH of Asp168. The side chain of Lys36, on the other hand, has several different orientations, and there are differences in the extent of infiltration of water molecules into this site (Fig. 5C). In molecules A and B, the Lys36 side chains turn back to hydrogen bond to the carbonyl oxygen of residue 37, whereas in molecule C it points in the opposite direction, hydrogen bonding to a water molecule that has entered the position vacated by the carboxyl group of the mutated Glu117. A lower occupancy water is at this position in molecule A, but there is none present in molecule B. Molecule C has a second water in this region, whereas its Oδ1 atom remains hydrogen-bonded to the main chain NH of Asp168. The side chain of Lys36, on the other hand, has several different orientations, and there are differences in the extent of infiltration of water molecules into this site (Fig. 5C). In molecules A and B, the Lys36 side chains turn back to hydrogen bond to the carbonyl oxygen of residue 37, whereas in molecule C it points in the opposite direction, hydrogen bonding to a water molecule that has entered the position vacated by the carboxyl group of the mutated Glu117. A lower occupancy water is at this position in molecule A, but there is none present in molecule B. Molecule C has a second water in this region, whereas its Oδ1 atom remains hydrogen-bonded to the main chain NH of Asp168. Concomitant with these changes there are reorientations of some of the hydrophobic side chains that cluster around the isopeptide bond-forming residues, notably Phe98, Phe98, and Ile56 (Fig. 6).

Overall, without the catalytic carboxyl group of Glu117 and consequent formation of the Lys36-Asn168 isopeptide bond, there is considerable flexibility and disturbance introduced to this part of the structure. The main chain conformation is not affected in any obvious way, which is not unexpected because the Lys and Asn residues are located on two adjacent β-strands, which are also held together by main chain C=O…NH bonds. Water molecules are, however, able to enter this region of the protein interior, and the hydrophobic core is disrupted.

**DISCUSSION**

The discovery in 2005 of the existence of pili on the surface of group A streptococcus (4) brought with it the recognition that the protein subunits that formed the polymeric pilus backbone corresponded to the Lancefield T antigens that had been described more than 50 years earlier (24). These antigens were known to be extremely robust in laboratory conditions and were given the designation “T” (trypsin-resistant) because of their strong resistance to trypsin digestion (24, 25). The basis for this unusual stability was not known, however, because nothing was known of either the structure or function of these antigens. The structure of Spy0128 (3), which corresponds to the T1 antigen of the M1 strain of GAS and which forms the backbone subunit of strain M1 GAS pili (4), gave the first indications of the source of this stability with the discovery of two...
intramolecular isopeptide bonds which join Lys and Asn residues in the core of the protein. Intramolecular bonds of this type were unprecedented, with the only parallel being the Lys-Asn isopeptide cross-links that impart extraordinary stability to the capsid of bacteriophage HK97 (12, 26).

We show here that the loss of the isopeptide bonds in Spy0128 severely compromises the stability of the protein. Mutagenesis, either of the Glu residues that are essential for the reaction that generates each isopeptide bond, or of the Asn residue from the Lys-Asn pair, results in proteins that lack one or both isopeptide bonds, as shown by mass spectrometry and x-ray crystallography. CD melting curves show that WT Spy0128 has high thermal stability, unfolding in a single cooperative transition at a melting temperature ($T_m$) of 85 °C. In contrast, the mutant proteins in which a single isopeptide bond is deleted by mutagenesis (E117A, E258A, and N168A) show biphasic unfolding, with transitions at $T_m$ of 50–60 °C and 80–85 °C. We interpret these results as indicating that the two domains unfold independently, with the domain that lacks an isopeptide bond (the N domain for E117A and N168A and the C domain for E258A) being some 30 °C less stable. The double mutant, which lacks both isopeptide bonds, is even more unstable and does not give a proper unfolding transition. The proteolytic stability is similarly compromised. The WT protein remained essentially intact under the conditions of our trypsin digestion experiment, whereas the single mutants were substantially degraded within a few hours, and the double mutant was degraded within minutes. Similar protease resistance is imparted by the isopeptide bonds in the B. cereus major pilin, BcpA (7).

CD measurements showed that the overall secondary structures were not affected by the loss of isopeptide bonds from the mutant proteins, and this was confirmed by the x-ray crystallographic analyses of two of the mutant proteins, E117A and N168A. The loss of stability of the mutant Spy0128 proteins can then be explained by several factors. First, the strong covalent cross-links clearly contribute to the structural integrity of the domain to which each one belongs. Their contribution to stability is likely to come not only from the strength of the covalent bond per se (similar to a normal peptide bond) but also from the strategic positioning of these bonds. Protein engineering studies have shown that introduced disulfide bonds, providing they are geometrically favorable, make a much greater contribution to stability when they join residues that are far apart in the linear amino acid sequence (2). This is because of their greater effect on the entropy of the unfolded polypeptide. In the case of Spy0128, the isopeptide bonds join the first and last β-strands of each domain, thereby having maximum entropic effect. Atomic force microscopy studies further show that, in Ig domain proteins such as protein L, titin, and tenascin, interactions between the two adjacent, parallel, terminal β-strands of their Ig domains impart the greatest resistance to mechanical unfolding (27, 28). This comes in large part from the backbone hydrogen-bonding network between the two strands, referred to as a “hydrogen bond clamp” (29), but would be strongly enhanced by the interstrand isopeptide bonds in Spy0128.

Not all of the loss of stability of the mutant proteins in this study can be attributed to the loss of the isopeptide bonds, however. Without the Lys-Asn cross-link, the charged side chain of Lys is left in an unfavorable environment, in the hydrophobic core, as seen in the crystal structure of Spy0128 E117A. Charged residues in the hydrophobic core are expected to be highly destabilizing, as is shown by the greatly reduced thermal stability when charged residues are introduced to the core of T4 lysozyme (30). The effects can be ameliorated by introduction of an opposite charge, as is shown for the staphylococcal ribonuclease RNAse Sα, for which the thermal instability caused by an isolated Asp in the hydrophobic core of the protein can be substantially relieved by introduction of a Lys side chain to balance the charge (31). In the structure of Spy0128 E117A, the mutation introduces internal flexibility and disruption to the hydrophobic core; the Lys$^{36}$ residue lacks significant stabilizing interactions with surrounding atoms, which could contribute to the loss of stability of the mutant protein. In contrast, in the N168A structure, a salt bridge is formed between Lys$^{36}$ and Glu$^{117}$, and there is virtually no disturbance to the surrounding structure. These structural differences are reflected in the differences observed in the proteolytic and thermal stability of these two mutant proteins. The two proteins lack the same N domain isopeptide bond, yet the E117A mutation is significantly more destabilizing than the N168A mutation as seen in the proteolytic digestion and thermal melting experiments.

We infer that the N168A mutant gives a better measure of the specific contribution of the isopeptide bond to stability. This mutant can be unfolded and refolded reversibly, like WT Spy0128, has similar CD spectra and a virtually identical structure, except for the loss of the isopeptide bond and formation of a Lys...Glu salt bridge, yet the N-terminal domain of this protein unfolds at a temperature that is ~25 °C lower.

Finally, the location of the isopeptide bonds in Spy0128 has an interesting relationship with the tandem arrangement of the backbone subunits in covalently polymerized pili. In each
domain of Spy0128, the internal cross-link occurs between the first and last $\beta$-strands. The last strand of the N-terminal domain continues on to become the first strand of the C-terminal domain, which in turn forms an isopeptide bond with the last strand of this domain. This final strand is then linked, via a sortase-mediated isopeptide bond to a lysine residue (Lys$^{161}$) of the next pilin subunit, located in a loop region just prior to the last strand of the N-terminal domain. As pointed out by Yeates and Clubb (32), this means that all the isopeptide bonds, both between and within subunits, are connected in an almost linear fashion along the entire pilus polymer. This covalent connectivity could be the main force-bearing feature of the pilus that must tolerate tensile forces while maintaining host cell attachment.

CONCLUSIONS

The intramolecular isopeptide bonds in Spy0128 seem likely to be important features of other Gram-positive pili and of many bacterial cell surface proteins that use CnaB- or CnaA-type domains (3). We have shown here that these covalent cross-links impart a high degree of thermodynamic stability, as indicated by thermal melting curves, as well as giving greatly increased resistance to proteolysis. Similar to disulfide bonds, the isopeptide bonds do not cause correct bonding but form via an intramolecular reaction in the folded structure. Consistent with this, Spy0128 remains correctly folded even without the isopeptide bonds but would clearly be much more prone to disruption by environmental or physical stress under in vivo conditions. This would be important for maintenance of the structural integrity of the streptococcal pilus and for that of other cell surface structures in which isopeptide bonds exist. There is also evidence from studies of the pili expressed by $B. cereus$ that these bonds are important for pilus assembly. BcpA, the major pilin of $B. cereus$, has three domains and three intramolecular isopeptide bonds (7). Deletion of the second of these, by mutation of the catalytic Glu$^{223}$ residue, abrogates pilus assembly (7). By analogy with the results described here for Spy0128, the most likely explanation is that the structure of BcpA is destabilized such that proper association with SrtD, the sortase that mediates assembly, is prevented.

Acknowledgments—We thank Drs. Juliet Gerrard and Grant Pearce (University of Canterbury, Christchurch, New Zealand) for help with CD measurements.

REFERENCES

1. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) FASEB J. 10, 75–83
2. Matsumura, M., Becktel, W. J., Levitt, M., and Matthews, B. W. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 6562–6566
3. Kang, H. J., Coulibaly, F., Chow, F., Proft, T., and Baker, E. N. (2007) Science 318, 1625–1628
4. Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A. G., Maggi, T., Taddei, A. R., Grandi, G., and Telford, J. L. (2005) Proc. Natl. Acad. Sci. U.S.A. 102, 15641–15646
5. Ton-That, H., and Schneewind, O. (2003) Mol. Microbiol. 50, 1429–1438
6. Telford, J. L., Barocchi, M. A., Margari, L., Rappuoli, R., and Grandi, G. (2006) Nat. Rev. Microbiol. 4, 509–519
7. Budzik, J. M., Marraffini, L. A., Souda, P., Whitelegge, J. P., Faull, K. F., and Schneewind, O. (2008) Proc. Natl. Acad. Sci. U.S.A. 105, 10215–10220
8. Pickart, C. M. (2001) Annu. Rev. Biochem. 70, 503–533
9. Greenberg, C. S., Birckbihler, P. J., and Rice, R. H. (1991) FASEB J. 5, 3071–3077
10. Ton-That, H., Marraffini, L. A., and Schneewind, O. (2004) Biochem. Biophys. Acta 1694, 269–278
11. Duda, R. L. (1998) Cell 94, 55–60
12. Wilkoff, W. R., Liljas, L., Duda, R. L., Tsuturuta, H., Hendrix, R. W., and Johnson, J. E. (2000) Science 289, 2129–2133
13. Krishnan, V., Gaspar, A. H., Ye, N., Mandlik, A., Ton-That, H., and Narayana, S. V. L. (2007) Structure 15, 893–903
14. Symersky, J., Patti, J. M., Carson, M., House-Pompeo, K., Teale, M., Moore, D., Jin, L., Schneider, A., DeLucas, L. J., Höök, M., and Narayana, S. V. L. (1997) Nat. Struct. Biol. 4, 833–838
15. Deivanayagam, C. C., Rich, R. L., Carson, M., Owens, R. T., Danthuluri, S., Bice, T., Höök, M., and Narayana, S. V. (2000) Structure 8, 67–78
16. Zong, Y., Xu, Y., Liang, X., Keene, D. R., Höök, A., Gurusiddappa, S., Höök, M., and Narayana, S. V. (2005) EMBO J. 24, 4224–4326
17. Studier, F. W. (2005) Protein Expr. Purif. 41, 207–234
18. Otwinowski, Z., and Minor, W. (1997) in Methods in Enzymology (Carter, J. W., and Sweet, R. M., eds) Vol. 276, pp. 307–326, Academic Press, New York
19. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. D. Biol. Crystallogr. 53, 290–255
20. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) J. Appl. Crystallogr. 40, 658–674
21. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Acta Crystallogr. D. Biol. Crystallogr. 57, 122–133
22. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D. Biol. Crystallogr. 60, 2126–2132
23. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
24. Lancefield, R. C., and Dole, V. P. (1946) J. Exp. Med. 84, 449–471
25. Schneewind, O., Jones, K. F., and Fischetti, V. A. (1990) J. Bacteriol. 172, 3310–3317
26. Dierkes, L. E., Peebles, C. L., Firek, B. A., Hendrix, R. W., and Duda, R. L. (2009) J. Virology 83, 2088–2098
27. Brockwell, D. J., Beddard, G. S., Pacci, E., West, D. K., Omlsted, P. D., Smith, D. A., and Radford, S. E. (2005) Biophys. J. 89, 506–519
28. Ng, S. P., Rounsevell, R. W., Steward, A., Geierhaas, C. D., Williams, P. M., Pacci, E., and Clarke, J. (2005) J. Mol. Biol. 350, 776–789
29. Forman, J. R., and Clarke, J. (2007) Curr. Opin. Struct. Biol. 17, 58–66
30. Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W., and Matthews, B. W. (1991) Biochemistry 30, 11521–11529
31. Trevino, S. R., Gokulan, K., Newsom, S., Thrulkill, R. L., Shaw, K. L., Mitkевич, V. A., Makarov, A. A., Sacchettini, J. C., Scholtz, J. M., and Pace, C. N. (2005) J. Mol. Biol. 354, 967–978
32. Yeates, T. O., and Clubb, R. T. (2007) Science 318, 1558–1559