

Science

AAAS

Stabilizing Isopeptide Bonds Revealed in Gram-Positive Bacterial Pilus Structure

Hae Joo Kang, *et al.*

Science **318**, 1625 (2007);

DOI: 10.1126/science.1145806

The following resources related to this article are available online at www.sciencemag.org (this information is current as of January 22, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/318/5856/1625>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/318/5856/1625/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/318/5856/1625#related-content>

This article **cites 23 articles**, 7 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/318/5856/1625#otherarticles>

This article has been **cited by** 1 article(s) on the ISI Web of Science.

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Stabilizing Isopeptide Bonds Revealed in Gram-Positive Bacterial Pilus Structure

Hae Joo Kang,^{1,2} Fasséli Coulibaly,^{1,2} Fiona Clow,^{1,3} Thomas Proft,^{1,3*} Edward N. Baker^{1,2*}

Many bacterial pathogens have long, slender pili through which they adhere to host cells. The crystal structure of the major pilin subunit from the Gram-positive human pathogen *Streptococcus pyogenes* at 2.2 angstroms resolution reveals an extended structure comprising two all- β domains. The molecules associate in columns through the crystal, with each carboxyl terminus adjacent to a conserved lysine of the next molecule. This lysine forms the isopeptide bonds that link the subunits in native pili, validating the relevance of the crystal assembly. Each subunit contains two lysine-asparagine isopeptide bonds generated by an intramolecular reaction, and we find evidence for similar isopeptide bonds in other cell surface proteins of Gram-positive bacteria. The present structure explains the strength and stability of such Gram-positive pili and could facilitate vaccine development.

Bacterial pili are filamentous structures that extend from the bacterial cell surface and mediate host cell adhesion, bacterial motility, and other critical aspects of colonization. The pili of pathogenic bacteria are also major virulence factors and important vaccine candidates. The best-characterized are the type I and type IV pili of Gram-negative organisms, for which considerable structural information exists on subunit structure and assembly (1–6). These pili are long (1 to 4 μm), thin (5 to 8 nm), and flexible, but are nonetheless very strong and can withstand extreme physical stresses.

By contrast, the pili on Gram-positive bacteria have mostly gone unrecognized until recently, probably because they are extremely thin (2 to 3 nm) and hard to see. Unlike Gram-negative pili, whose subunits associate via noncovalent interactions, Gram-positive pili are assembled by bacterially encoded transpeptidase enzymes called sortases. These enzymes recognize specific sequence motifs in the pilin subunits, elongate the pilus oligomer by progressive addition of subunits joined by intermolecular isopeptide bonds, and then tether the entire assembly to the cell wall peptidoglycan (7–9). The pili thus consist of multiple, covalently bonded copies of a single backbone pilin, to which can be added a few accessory proteins.

Streptococcus pyogenes [group A *Streptococcus* (GAS)] infects the human throat and skin, causing common infections such as sore throat and tonsillitis, as well as severe invasive illnesses such as necrotizing fasciitis, rheumatic fever, and streptococcal toxic shock syndrome (10). Thin pili, ~ 2 nm wide and >1 μm long, have been revealed by electron microscopy (11) and were shown to be essential for adhesion to

human tonsil and skin cells (12) as well as promising vaccine candidates against virulent GAS bacteria (11). The pilus-forming proteins are encoded in a small gene cluster within a pathogenicity island known as the FCT (fibronectin-binding, collagen-binding T antigen) region. In the *S. pyogenes* M1 strain SF370, *spy0128* en-

codes the backbone pilin, *spy0129* the sortase C1, and *spy0125* and *spy0130* two pilin-associated proteins (11). The backbone pilin subunits are Lancefield T antigens (13, 14), named for their antigenicity and their extreme resistance to trypsin (T) digestion.

To understand pilus stability and assembly in Gram-positive organisms, we expressed the backbone pilin protein Spy0128 from an M1 strain of *S. pyogenes*. This 340-residue protein has a sortase recognition motif, Glu-Val-Pro-Thr-Gly, at residues 308 to 312. Constructs comprising residues 18 to 311 and 18 to 308 were prepared. We obtained excellent crystals for the latter (15) and solved its crystal structure at 2.2 Å resolution ($R = 20.3\%$, $R_{\text{free}} = 26.4\%$) (table S1).

The Spy0128 monomer has an elongated two-domain structure, with length 98 Å and width 20 to 30 Å (Fig. 1A). Both domains have irregular all- β structures that are modified variants of the immunoglobulin fold (Fig. 1B). The N-terminal domain, residues 18 to 171, forms a β sandwich in which the strands in one β sheet (green in Fig. 1) are progressively extended such that the upper portion of this β sheet, at the top of the domain as shown in Fig. 1, is relatively exposed. The C-terminal domain,

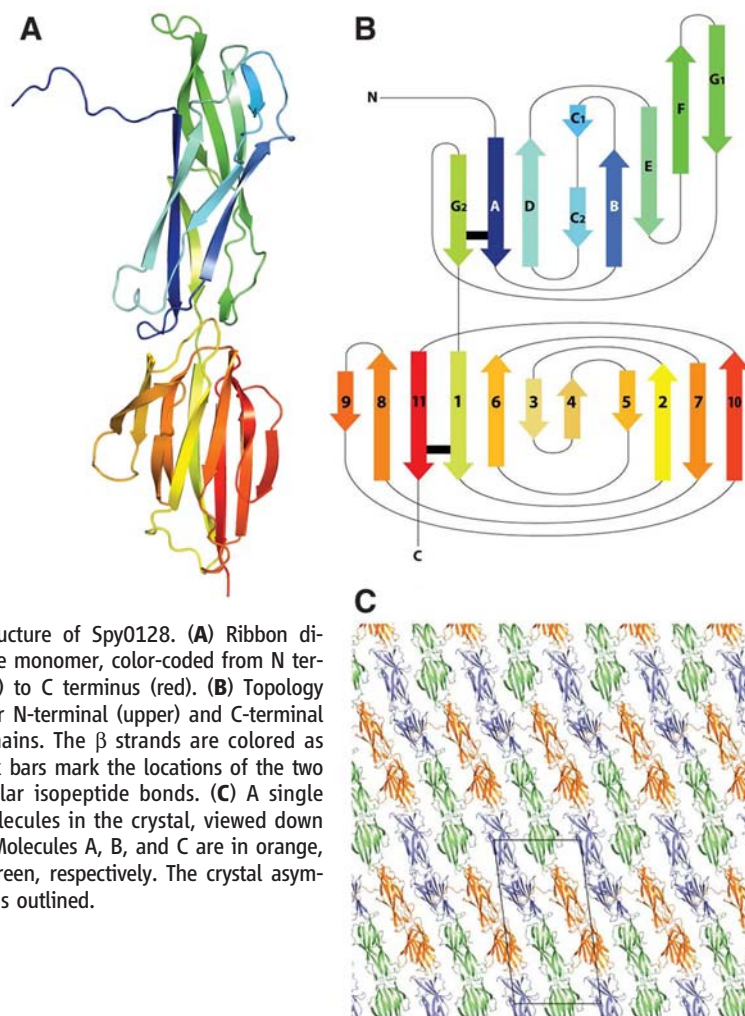


Fig. 1. Structure of Spy0128. (A) Ribbon diagram of the monomer, color-coded from N terminus (blue) to C terminus (red). (B) Topology diagrams for N-terminal (upper) and C-terminal (lower) domains. The β strands are colored as in (A). Black bars mark the locations of the two intramolecular isopeptide bonds. (C) A single layer of molecules in the crystal, viewed down the b axis. Molecules A, B, and C are in orange, blue, and green, respectively. The crystal asymmetric unit is outlined.

¹Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland 1010, New Zealand. ²School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand. ³School of Medical Sciences, University of Auckland, Auckland 1023, New Zealand.

*To whom correspondence should be addressed. E-mail: ted.baker@auckland.ac.nz (E.N.B.); t.proft@auckland.ac.nz (T.P.)

residues 173 to 307, comprises 11 β strands. Its core is a β sandwich in which a five-stranded β sheet packs against a four-stranded β sheet. A prominent β ribbon (strands $\beta 3$ and $\beta 4$) extends the first sheet to seven strands and provides a wide loop at the base of the domain. Overall, the domain is wedge-shaped with a broad base and a narrower top where it joins to the N domain. The two domains are intimately associated, with only one residue, Ser¹⁷², separating the final β strand of the N domain from the first of the C domain. The interface between domains is mostly hydrophobic and buries $\sim 1200 \text{ \AA}^2$ of surface area.

The crystal asymmetric unit contains three independent Spy0128 molecules that generate columns of molecules extending through the crystal (Fig. 1C). This arrangement, found also in another crystal form (15), provides a compelling model for the assembly of GAS pili. Successive molecules stack head-to-tail, related by an approximate 3_1 helical screw along their long axis. Each interface, between the N domain of one molecule and the C domain of the next (Fig. 2), buries $\sim 850 \text{ \AA}^2$ of solvent-accessible surface with a shape complementarity of 0.72, comparable with other protein oligomerization interfaces (16). There is very little lateral interaction between columns of molecules in the crystal.

The head-to-tail packing means that Phe³⁰⁷, which closely precedes the sortase recognition motif in Spy0128, packs against the exposed face of the N-domain β sheet (Fig. 2B). Sortase action cleaves the Thr³¹¹-Gly³¹² bond, after which isopeptide bond formation between the new C terminus and a Lys residue covalently links adjacent pilin subunits (7, 8). Five invariant lysines (Fig. 2C) are potential candidates for this intermolecular linkage. Of these, only Lys¹⁶¹, near the top of the N domain and 11 to 13 \AA below Phe³⁰⁷ of the next molecule in the column, is a viable candidate for generating an elongated pilus. We used mass spectrometry of pilus fractions from *S. pyogenes* to show that Lys¹⁶¹ is indeed the essential lysine involved in oligomerization (fig. S3). This finding strongly supports the biological relevance of the assembly seen in all crystal forms. Residues 308 to 311 would continue below Phe³⁰⁷, packing against a highly sequence-conserved region of the β sheet (Fig. 2B) and allowing isopeptide bond formation between the Thr³¹¹ carboxyl and Lys¹⁶¹ N ζ of the next molecule.

Intermolecular isopeptide bonds are known in other contexts besides the sortase-generated isopeptide bonds of Gram-positive pili. In ubiquitination, specific lysine residues of a target protein are covalently linked by ubiquitin ligases to the terminal carboxylate of ubiquitin (17). In transglutamination, enzyme-catalyzed isopeptide bond formation occurs between Gln and Lys side chains (18), as in the cross-linking of fibrin subunits, catalyzed by factor XIII (19). A rare example of self-generated isopeptide bonds between Asn and Lys residues occurs in the bacteriophage HK97, where capsid subunits are covalently cross-linked to form interlocked circular rings

that give extraordinary stability (20, 21). No examples of intramolecular isopeptide bonds have been reported, however.

In this context we were surprised to observe two intramolecular isopeptide bonds within the pilin subunit, one in each domain. Formed by covalent bonding between lysine and asparagine side chains (Lys³⁶-Asn¹⁶⁸ in the N domain; Lys¹⁷⁹-Asn³⁰³ in the C domain), these are each indicated by continuous electron density extending through the lysine ϵ -amino group into the δ -carboxamide group of asparagine (Fig. 3A). Mass spectrometry provided independent confirmation (15). The protein molecular mass was consistent with the loss of two NH_3 units through isopeptide bond formation (table S2), and proteolytic digestion and peptide mapping gave cleavage products containing nonconsecutive sequences (figs. S1 and S2). These mapped to peptides surrounding both isopeptide bonds.

These bonds appear, as in HK97, to be self-generated. An essential Glu residue is associated with each bond, forming hydrogen bonds to the isopeptide C=O and NH groups (Fig. 3A). The hydrogen bonding implies that both glutamic acids, Glu¹¹⁷ and Glu²⁵⁸, are protonated. In each case, the Lys, Asn, and Glu residues are surrounded by a cluster of aromatic residues (Fig.

3C), which would favor elevation of the pK_a of the glutamic acid and reduction of the pK_a of the lysine ϵ -amino group. In the N domain, the isopeptide moiety sits over the aromatic ring plane of Phe⁵², and Glu¹¹⁷ is surrounded by Phe⁵⁴, Tyr¹²⁸, and Phe¹⁶⁶. Similar roles are played by Phe¹⁹², Phe¹⁹⁴, Tyr²⁶¹, and Phe³⁰¹ for the C-terminal isopeptide. A plausible mechanism for isopeptide bond formation, first suggested for HK97 (21), is that the protonated Glu polarizes the C=O bond of the Asn side chain, inducing positive charge on C γ . Nucleophilic attack on C γ by the unprotonated Lys ϵ -amino group then generates the isopeptide bond.

We tested the importance of the Glu residues for isopeptide formation by mutating Glu¹¹⁷ and Glu²⁵⁸ to alanine, creating proteins E117A and E258A. Mass spectrometry showed the loss of one isopeptide bond from each mutant (table S2), and crystallographic analysis of E117A confirmed that the N-domain isopeptide was not formed when Glu¹¹⁷ was mutated (fig. S4). Both mutants also showed greatly increased susceptibility to proteolysis, indicating the stabilizing effect of these cross-links (fig. S5).

Sequence comparisons (fig. S6) suggested that the isopeptide bonds may be a conserved feature of the pili of all GAS. Despite low over-

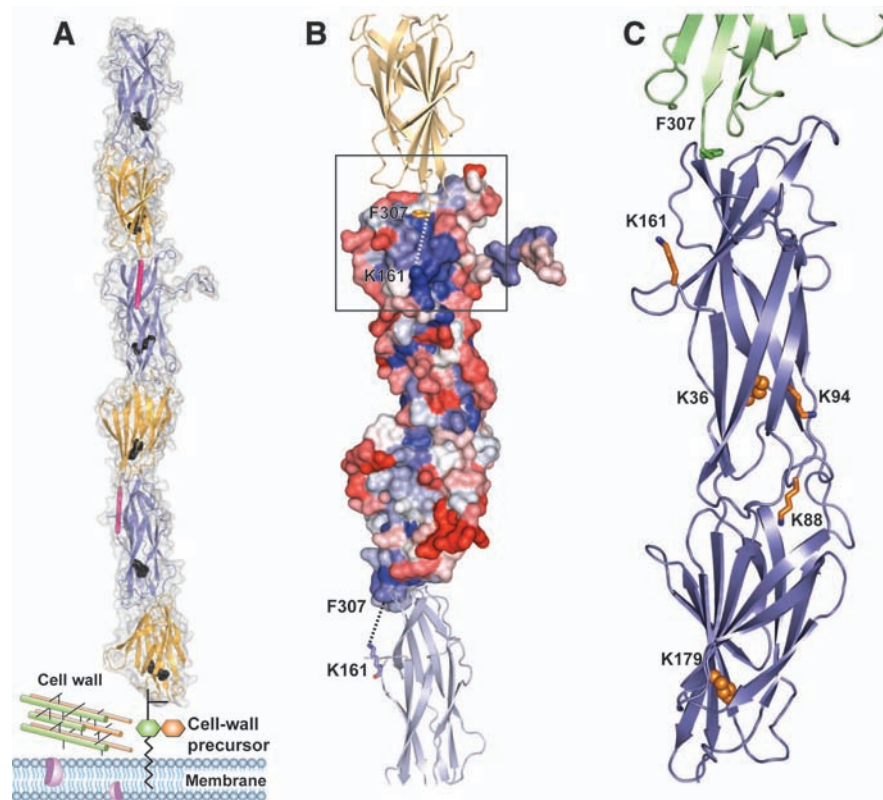


Fig. 2. Assembly of *S. pyogenes* pili. (A) Column of Spy0128 molecules in which the N domain (blue) of one molecule packs against the C domain (gold) of the next. Black bars indicate the intramolecular isopeptide bonds and red bars the proposed intermolecular links. (B) Pilus assembly in which one molecule is shown in surface representation, colored by sequence conservation across GAS major pili (table S2) from dark blue (highly conserved) to dark red (highly variable). A broken line shows the location of the sortase-mediated intermolecular isopeptide bond, joining the C terminus to the invariant Lys¹⁶¹. F, Phe; K, Lys. (C) Location of all conserved lysines in Spy0128.

all sequence identity in Spy0128 alleles, the Lys, Asn, and Glu residues of the isopeptide bonds are strictly conserved, as are five of the eight aromatic residues surrounding them. The other aromatics are replaced only by hydrophobic residues. The isopeptide bonds are strategically located in each domain (just before the interdomain connection and the sortase recognition

motif, respectively), tying together the first and last β strands (Fig. 1B). Sequence similarities with the major pilins from other Gram-positive bacteria are too low to determine whether isopeptide bonds are a common feature, but a conserved Asn precedes the sortase motif by 5 to 8 residues in all sequences we have examined (fig. S7), and conserved Lys and Glu residues can also be traced.

We also found evidence for intramolecular isopeptide bonds in other cell surface proteins. The C-terminal domain of the pilin-associated Cpa (GAS collagen-binding protein), encoded by *spy0125*, is homologous with the C domain of Spy0128, with residues involved in the C-terminal isopeptide bond (Lys, Asn, Glu, and three Phe) invariant across all 14 Cpa sequences in the current sequence database (fig. S8). Examination of the recently released structure of a minor pilin, GBS52 from *Streptococcus agalactiae* (22), reveals an unrecognized Lys-Asn isopeptide bond like those in Spy0128 (fig. S9). We then searched the Protein Data Bank, using a Lys-Asn-Glu/Asp structural template, and identified the collagen-binding adhesin Cna from *Staphylococcus aureus* (23, 24) as also having previously unrecognized isopeptide bonds in its A and B domains (figs. S9 and S10). Further sequence searches showed many instances of these domains containing predicted isopeptide bond-forming residues in the same locations (Fig. 4), all from Gram-positive organisms and all (where functionally characterized) cell surface adhesion proteins.

The isopeptide bonds we have found in GAS pili and other Gram-positive adhesins provide a striking parallel with the disulfide bridges found in Gram-negative pilins and adhesins (1, 5), which are important for pilus assembly and substrate binding. We hypothesize that in Gram-positive organisms, which lack the disulfide bond formation machinery of Gram-negative bacteria, intramolecular isopeptide bonds may provide an alternative mode of stabilization for cell surface proteins involved in host pathogenesis.

Our results provide a model for the assembly of *S. pyogenes* pili, in which self-generated intramolecular isopeptide bonds complement the sortase-catalyzed intermolecular bonds. The long, thin GAS pili are only ~2 nm (one molecule) thick (11) but typically >100 molecules long, and we infer that these bonds play a crit-

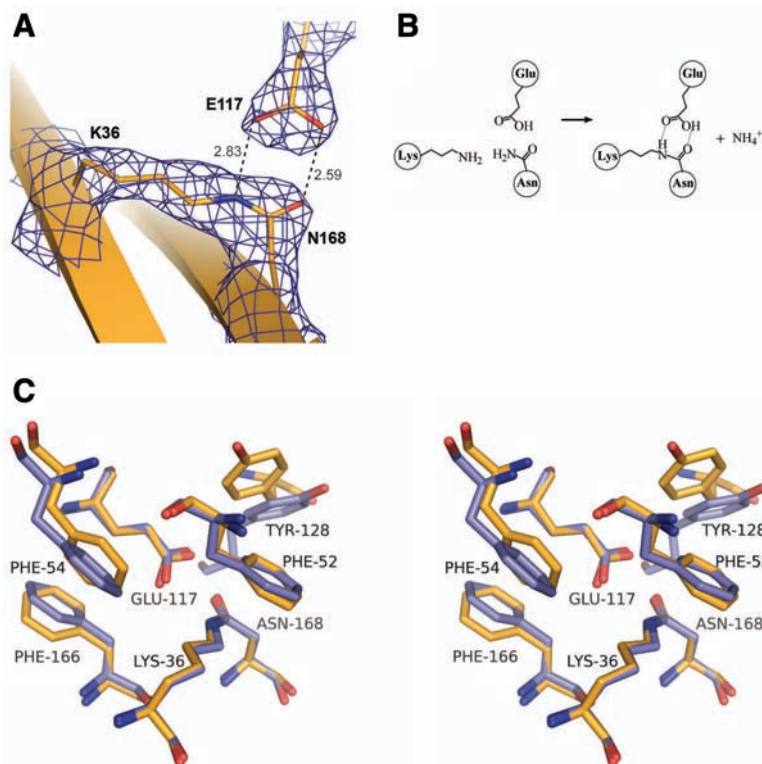
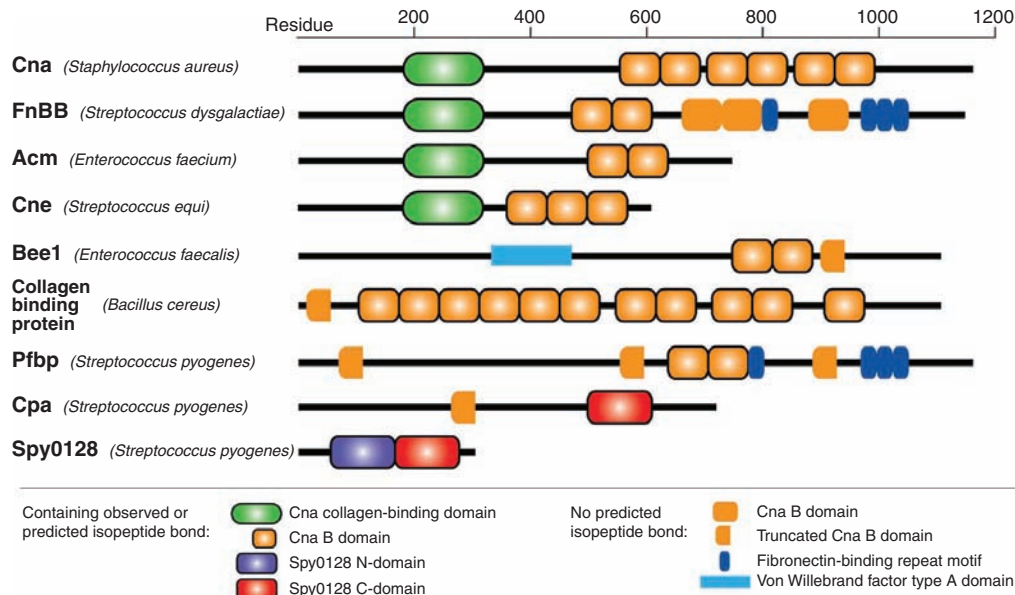


Fig. 3. Isopeptide bond formation in Spy0128. **(A)** The Lys-Asn (K-N) isopeptide bond in the N domain, shown in stick mode, in its $F_{obs} - F_{calc}$ omit density, contoured at 1.5σ . Hydrogen bonds with the catalytic Glu¹¹⁷ (E117) are shown with broken lines, with distances in Å. **(B)** The chemical reaction for isopeptide bond formation. **(C)** Stereo view of a superposition of the N-terminal (blue) and C-terminal (orange) isopeptide bonds, with the conserved aromatic residues surrounding them.

Fig. 4. Domain organization of proteins containing observed or predicted isopeptide bonds. Only proteins with known functions are shown. Domains with the conserved Lys-Asn-Glu of Spy0128 and CnaB or Lys-Asn-Asp of CnaA are highlighted by black outlines.



ical role in maintaining pilus integrity in the face of severe mechanical and chemical stress while bound to host cells. GAS pili show considerable antigenic variation, indicating an important role in virulence, and the pilin subunits are T antigens that are used for serotyping (13). The presence of several conserved regions on a highly variable background (Fig. 2B) suggests that the structure could help provide an effective pilus-based vaccine against GAS.

References and Notes

1. D. Choudhury *et al.*, *Science* **285**, 1061 (1999).
2. E. Hahn *et al.*, *J. Mol. Biol.* **323**, 845 (2002).
3. H. Remaut, G. Waksman, *Curr. Opin. Struct. Biol.* **14**, 161 (2004).
4. H. E. Parge *et al.*, *Nature* **378**, 32 (1995).
5. L. Craig, M. E. Pique, J. A. Tainer, *Nat. Rev. Microbiol.* **2**, 363 (2004).
6. L. Craig *et al.*, *Mol. Cell* **23**, 651 (2006).

7. H. Ton-That, O. Schneewind, *Mol. Microbiol.* **50**, 1429 (2003).
8. J. L. Telford, M. A. Barocchi, I. Margarit, R. Rappuoli, G. Grandi, *Nat. Rev. Microbiol.* **4**, 509 (2006).
9. H. Ton-That, L. A. Marraffini, O. Schneewind, *Biochim. Biophys. Acta* **1694**, 269 (2004).
10. M. W. Cunningham, *Clin. Microbiol. Rev.* **13**, 470 (2000).
11. M. Mora *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15641 (2005).
12. E. L. Abbot *et al.*, *Cell. Microbiol.* **9**, 1822 (2007).
13. R. C. Lancefield, *J. Exp. Med.* **71**, 521 (1940).
14. R. C. Lancefield, V. P. Dole, *J. Exp. Med.* **84**, 449 (1946).
15. See supporting material on Science Online.
16. M. C. Lawrence, P. M. Colman, *J. Mol. Biol.* **234**, 946 (1993).
17. C. M. Pickart, *Annu. Rev. Biochem.* **70**, 503 (2001).
18. C. S. Greenberg, P. J. Birckbichler, R. H. Rice, *FASEB J.* **5**, 3071 (1991).
19. J. W. Weisel, *Adv. Protein Chem.* **70**, 247 (2005).
20. W. R. Wikoff *et al.*, *Science* **289**, 2129 (2000).
21. C. Helgstrand *et al.*, *J. Mol. Biol.* **334**, 885 (2003).
22. V. Krishnan *et al.*, *Structure* **15**, 893 (2007).

23. C. C. S. Deivanayagam *et al.*, *Structure* **8**, 67 (2000).
24. Y. Zong *et al.*, *EMBO J.* **24**, 4224 (2005).
25. Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank with accession number 3B2M. We thank T. Caradoc-Davies for help with refinement and M. Middleditch for help with mass spectrometry. Supported by the Health Research Council (HRC) of New Zealand, the Foundation for Research, Science and Technology (FRST) of New Zealand, and the Maurice Wilkins Centre for Molecular Biodiscovery. H.J.K. received an FRST Bright Futures Scholarship, and T.P. is an HRC Hercus Fellow.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5856/PAGE/DC1
 Materials and Methods
 SOM Text
 Figs. S1 to S13
 Tables S1 to S4
 References

29 May 2007; accepted 23 October 2007
 10.1126/science.1145806

DUBA: A Deubiquitinase That Regulates Type I Interferon Production

Nobuhiko Kayagaki,¹ Qui Phung,² Salina Chan,¹ Ruchir Chaudhari,¹ Casey Quan,¹ Karen M. O'Rourke,¹ Michael Eby,¹ Eric Pietras,³ Genhong Cheng,³ J. Fernando Bazan,⁴ Zemin Zhang,⁵ David Arnott,² Vishva M. Dixit^{1*}

Production of type I interferon (IFN-I) is a critical host defense triggered by pattern-recognition receptors (PRRs) of the innate immune system. Deubiquitinating enzyme A (DUBA), an ovarian tumor domain-containing deubiquitinating enzyme, was discovered in a small interfering RNA-based screen as a regulator of IFN-I production. Reduction of DUBA augmented the PRR-induced IFN-I response, whereas ectopic expression of DUBA had the converse effect. DUBA bound tumor necrosis factor receptor-associated factor 3 (TRAF3), an adaptor protein essential for the IFN-I response. TRAF3 is an E3 ubiquitin ligase that preferentially assembled lysine-63-linked polyubiquitin chains. DUBA selectively cleaved the lysine-63-linked polyubiquitin chains on TRAF3, resulting in its dissociation from the downstream signaling complex containing TANK-binding kinase 1. A discrete ubiquitin interaction motif within DUBA was required for efficient deubiquitination of TRAF3 and optimal suppression of IFN-I. Our data identify DUBA as a negative regulator of innate immune responses.

Innate immune responses are initiated when host cellular PRRs encounter pathogen-associated molecular patterns (PAMPs) (1). Double- and single-stranded RNAs are virus-

derived PAMPs that trigger the intracellular PRRs Toll-like receptor 3 (TLR3), retinoic acid-inducible protein 1 (RIG-I), and melanoma differentiation-associated gene 5 (MDA5) (2–4).

Activation of these intracellular sensors leads to the recruitment of adaptor proteins for interferon- α (IFN- α) and IFN- β production. Toll-interleukin 1 receptor domain-containing adaptor inducing IFN- β (TRIF) interacts with TLR3, whereas IFN- β promoter stimulator 1 [(IPS-1), also called Cardif, MAVS, and VISA] is recruited by RIG-I and MDA5. These adaptors mediate the assembly of a signaling complex composed of the ubiquitin ligase TRAF3 and the kinases TANK-binding kinase 1 (TBK1) and inhibitor of nuclear factor κ B kinase ϵ [(IKK ϵ), also called IKK i (1, 5–7)]. This complex activates the downstream transcription factors, IFN regulatory factors 3 and 7 (IRF3 and IRF7), to switch on IFN-I expression, which is an essential aspect

¹Department of Physiological Chemistry, Genentech, South San Francisco, CA 94080, USA. ²Department of Protein Chemistry, Genentech, South San Francisco, CA 94080, USA. ³Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095, USA. ⁴Department of Protein Engineering, Genentech, South San Francisco, CA 94080, USA. ⁵Department of Bioinformatics, Genentech, South San Francisco, CA 94080, USA.

*To whom correspondence should be addressed. E-mail: dixit@gene.com

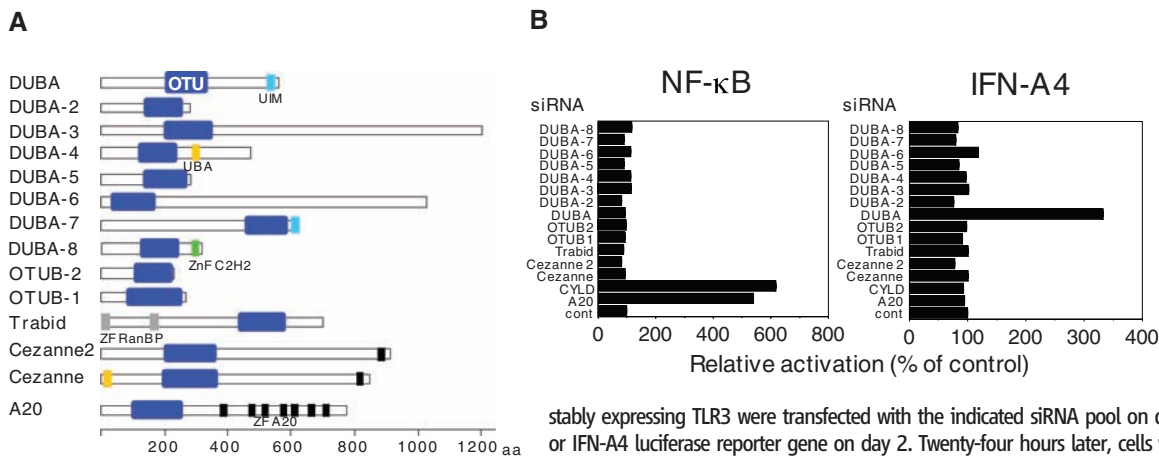


Fig. 1. An siRNA-based screen for OTU DUB family members. (A) Schematic of OTU family members. Gene accession numbers are listed in table S1. ZnF C2H2, zinc finger domain (C2H2-type); ZF RanBP, zinc finger domain (Ran-binding protein and others); ZF A20, zinc finger domain (A20-like); aa, amino acids. (B) HEK293 cells

stably expressing TLR3 were transfected with the indicated siRNA pool on day 0 and then with a NF- κ B or IFN-A4 luciferase reporter gene on day 2. Twenty-four hours later, cells were stimulated with poly(I:C) (20 μ g/ml) for 24 hours and reporter activation was measured. cont, control.