



A Conserved Class II Type Thioester Domain-Containing Adhesin Is Required for Efficient Conjugation in *Bacillus subtilis*

 César Gago-Córdoba,^a
 Jorge Val-Calvo,^a
 David Abia,^a
 Alberto Díaz-Talavera,^a
 Andrés Miguel-Arribas,^a
 Rocío Aguilar Suárez,^b
 Jan Maarten van Dijk,^b
 Ling Juan Wu,^c
 Wilfried J. J. Meijer^a

^aCentro de Biología Molecular Severo Ochoa (CSIC-UAM), Universidad Autónoma, Madrid, Spain

^bUniversity of Groningen, University Medical Center Groningen, Groningen, The Netherlands

^cCentre for Bacterial Cell Biology, Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

ABSTRACT Conjugation, the process by which a DNA element is transferred from a donor to a recipient cell, is the main horizontal gene transfer route responsible for the spread of antibiotic resistance and virulence genes. Contact between a donor and a recipient cell is a prerequisite for conjugation, because conjugative DNA is transferred into the recipient via a channel connecting the two cells. Conjugative elements encode proteins dedicated to facilitating the recognition and attachment to recipient cells, also known as mating pair formation. A subgroup of the conjugative elements is able to mediate efficient conjugation during planktonic growth, and mechanisms facilitating mating pair formation will be particularly important in these cases. Conjugative elements of Gram-negative bacteria encode conjugative pili, also known as sex pili, some of which are retractile. Far less is known about mechanisms that promote mating pair formation in Gram-positive bacteria. The conjugative plasmid pLS20 of the Gram-positive bacterium *Bacillus subtilis* allows efficient conjugation in liquid medium. Here, we report the identification of an adhesin gene in the pLS20 conjugation operon. The N-terminal region of the adhesin contains a class II type thioester domain (TED) that is essential for efficient conjugation, particularly in liquid medium. We show that TED-containing adhesins are widely conserved in Gram-positive bacteria, including pathogens where they often play crucial roles in pathogenesis. Our study is the first to demonstrate the involvement of a class II type TED-containing adhesin in conjugation.

IMPORTANCE Bacterial resistance to antibiotics has become a serious health care problem. The spread of antibiotic resistance genes between bacteria of the same or different species is often mediated by a process named conjugation, where a donor cell transfers DNA to a recipient cell through a connecting channel. The first step in conjugation is recognition and attachment of the donor to a recipient cell. Little is known about this first step, particularly in Gram-positive bacteria. Here, we show that the conjugative plasmid pLS20 of *Bacillus subtilis* encodes an adhesin protein that is essential for effective conjugation. This adhesin protein has a structural organization similar to adhesins produced by other Gram-positive bacteria, including major pathogens, where the adhesins serve in attachment to host tissues during colonization and infection. Our findings may thus also open novel avenues to design drugs that inhibit the spread of antibiotic resistance by blocking the first recipient-attachment step in conjugation.

KEYWORDS adhesion molecules, antibiotic resistance, conjugation, Gram-positive bacteria, mating, plasmids, thioester domain

Bacteria exchange DNA at large scale by different routes which are collectively called horizontal gene transfer (HGT) (for review, see references 1 to 4). HGT shapes

Citation Gago-Córdoba C, Val-Calvo J, Abia D, Díaz-Talavera A, Miguel-Arribas A, Aguilar Suárez R, van Dijk JM, Wu LJ, Meijer WJJ. 2021. A conserved class II type thioester domain-containing adhesin is required for efficient conjugation in *Bacillus subtilis*. mBio 12:e00104-21. <https://doi.org/10.1128/mBio.00104-21>.

Editor Rino Rappuoli, GSK Vaccines

Copyright © 2021 Gago-Córdoba et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Wilfried J. J. Meijer, wmeijer@cbm.csic.es.

Received 13 January 2021

Accepted 5 February 2021

Published 16 March 2021

the genetic content and plays a major role in the evolution of bacteria. The downside of HGT is that it contributes importantly to the emergence and dissemination of antibiotic resistance (5), which is one of today's major health care problems (6).

Conjugation, the process by which a conjugative DNA element is transferred from a donor to a recipient cell via a connecting channel, is the principal HGT route that is responsible for the spread of antibiotic resistance (5, 7, 8). Conjugative elements that are inserted in a bacterial genome are called integrative and conjugative elements (ICEs), and those present on plasmids are named conjugative plasmids. The basic principles of conjugation are conserved among ICEs and conjugative plasmids present in Gram-negative (G^-) or Gram-positive (G^+) bacteria. Conjugation occurs when cells are in close contact with each other, for instance in biofilms. However, a subset of conjugative elements is able to mediate conjugation with high efficiency also during planktonic growth. A prerequisite for conjugation is that donor cells recognize and contact a recipient cell in a process named mating pair formation (MPF). Particularly during conjugation in liquid medium, the presence of adhesive molecules or organelles is important for establishing contact between donor and recipient cells. Sophisticated type IV secretion systems (T4SS) are known to connect the donor and recipient cells, through which a copy of the conjugative DNA element is transported. All conjugative T4SS of G^- bacteria involve pili, also known as sex pili, which extend from the donor surface into the extracellular space and are involved in recipient cell recognition. Among the best studied are the pili encoded by the *Agrobacterium tumefaciens* pTi and the *Escherichia coli* F plasmids, named T- and F-pilus, respectively. The biogenesis and function of the T-pilus have recently been reviewed (9, 10). The T-pilus is formed as an extension at the surface of the T4SS and consists of multiple VirB2 shaft subunits and a VirB5 pilus tip adhesin protein. F-pili are long, flexible, and occasionally retractile. Their synthesis is more complex than that of T-pili (for review, see references 11 to 13).

Much less is known about how G^+ donor cells attach to recipient cells, but it seems that T4SS of G^+ bacteria do not develop pili. The only T4SS of a G^+ system for which recipient attachment has been studied in detail is PrgB encoded by the *Enterococcus faecalis* conjugative plasmid pCF10 (for review, see reference 14). Recently, the structure of the PrgB adhesin domain was shown to resemble the lectin-like fold in general and the multimodal *Streptococcus* sp. AgII/II, SspB, and SpaP adhesins in particular (15).

Conjugative plasmid pLS20 from the G^+ bacterium *Bacillus subtilis* and its derivative pLS20cat, which contains a chloramphenicol resistance gene, conjugate efficiently in both solid and liquid media (16–18). We therefore reasoned that pLS20 would encode a protein(s) that allows efficient MPF, especially in liquid media. pLS20 contains a large conjugation operon encompassing genes 28 to 74 (according to our gene annotation [19]). Expression of the conjugation operon is controlled by a strong promoter, P_c , whose activity is regulated by proteins encoded by genes 25 to 27 (19, 20). Here, we show that in the absence of pLS20cat gene 34 conjugation was severely affected in liquid media and moderately in solid media. Gene 34 encodes a 778-residue-long protein with a predicted N-terminal signal peptide. We present evidence that protein p34 is an adhesin that contains a class II type thioester domain (TED). A point mutation predicted to prevent formation of the thioester bond rendered the protein inactive. Sequence similarity and modeling strongly indicate that the TED is followed by CnaB-type structures with intramolecular isopeptide bonds that function potentially as a stalk to position the TED-type adhesin away from the cell surface. Therefore, protein p34 belongs to a family of G^+ bacterial proteins that are characterized by intramolecular cross-links in structurally conserved thioester, isopeptide, and ester domains and are hence named TIE proteins. Correspondingly, we have named p34 TIE_{pLS20}. TIE proteins have previously been shown to play important roles in the virulence of various pathogenic bacteria. Our results show, for the first time, that a TED-containing adhesin also plays an important role in plasmid conjugation.

RESULTS

pLS20cat gene 34 is required for efficient conjugation in liquid medium. In our recent study on exclusion proteins, we screened all pLS20cat genes for candidates encoding putative surface proteins (21). Besides the exclusion gene *ses_{pLS20}* (gene 29), gene 34 was also predicted to encode a surface protein. In the two plasmid sequences deposited in the NCBI database (accession numbers [AB615352.1](#) and [NC_015148.1](#)), pLS20cat gene 34 is annotated as a gene encoding a 753-residue protein ([BAJ76911.1](#) and [YP_004243498.1](#), respectively). However, the proposed GTG start codon is not preceded by a good ribosomal binding site. Importantly, the reading frame could be extended at the 5' end for 25 codons, where another putative GTG start codon is preceded by a ribosome binding site (RBS) (AAAGGGG-8 bp-GTG). The deduced sequence of the longer version of protein p34 (778 residues) contains an N-terminal signal peptide, indicating that the protein is exported from the cytoplasm, which is consistent with p34 sharing (low-level) similarity to a number of adhesins (see below). Based on this, we hypothesized that p34 could be an adhesin with a role in MPF. As a first strategy to test this idea, we constructed a derivative of pLS20cat, named pLS20catΔ34, containing a large in-frame deletion in gene 34. Strain CG164 harboring pLS20catΔ34 was employed as a donor in conjugation experiments using standard conditions to determine conjugation efficiency in liquid medium. In parallel, conjugation experiments were performed using as donor strain PKS11 harboring wild-type pLS20cat. In agreement with previous results, conjugation efficiencies in the range of 1×10^{-3} (calculated as transconjugants per donor) were obtained for pLS20cat (Fig. 1A). However, about ~1,000-fold-lower conjugation efficiencies were obtained for pLS20catΔ34, demonstrating that gene 34 is required for efficient conjugation in liquid medium. In these experiments, the spectinomycin-resistant strain PS110 was used as recipient. To verify that the antibiotic resistance marker had no effect on conjugation efficiencies, we repeated these experiments using as recipient the erythromycin-resistant strain PKS7. Similarly reduced conjugation efficiencies were obtained when gene 34 was disrupted regardless of the antibiotic marker used (Fig. 1A).

Next, we investigated whether gene 34 was also required for efficient conjugation on solid medium. Importantly, although inactivation of gene 34 also resulted in lower conjugation efficiencies on solid medium, the effect was much smaller than that observed in liquid medium (Fig. 1B). Compared to pLS20cat, the conjugation efficiencies of pLS20catΔ34 were about 1,000- and 15-fold lower in liquid and solid media, respectively. Together, these results demonstrate that pLS20cat gene 34 is required for efficient conjugation, particularly in liquid medium.

Ectopic expression of pLS20cat gene 34 in donor cells, but not in recipient cells, restores efficient conjugation of pLS20catΔ34. To test whether ectopic expression of gene 34 could complement the deletion of gene 34 in pLS20cat, we constructed *B. subtilis* strain CG157 (*amyE::P_{hyspank}-34*) that allowed conditional expression of pLS20cat gene 34 from the chromosome. Next, pLS20catΔ34 was introduced into CG157 to generate donor strain CG159 (*amyE::P_{hyspank}-34*, pLS20catΔ34). Low and high conjugation levels were obtained in liquid conjugation experiments in the absence or presence of a 1 mM concentration of the inducer IPTG (isopropyl-β-D-1-thiogalactopyranoside), respectively (Fig. 2A), demonstrating that ectopic expression of gene 34 in donor cells restored efficient conjugation of plasmid pLS20catΔ34. Possible effects of ectopic expression of gene 34 in the recipient cells were also tested, by mating PKS11 (pLS20cat) or CG164 (pLS20catΔ34) with CG157 (*amyE::P_{hyspank}-34*) recipient cells growing in the absence or presence of a 1 mM concentration of the inducer IPTG. Ectopic expression of gene 34 in recipient cells did not result in increased conjugation efficiency of either pLS20cat or pLS20catΔ34.

In silico analyses of the deduced protein p34 sequence suggests that it is an adhesin. *In silico* approaches were used to gain insights into features of protein p34 that might explain its importance in conjugation in liquid medium. The deduced sequence of protein p34 was subjected to the TMHMM 2.0 server to predict transmembrane helices and to the signal peptide-screening server SignalP v 5.0 (22, 23).

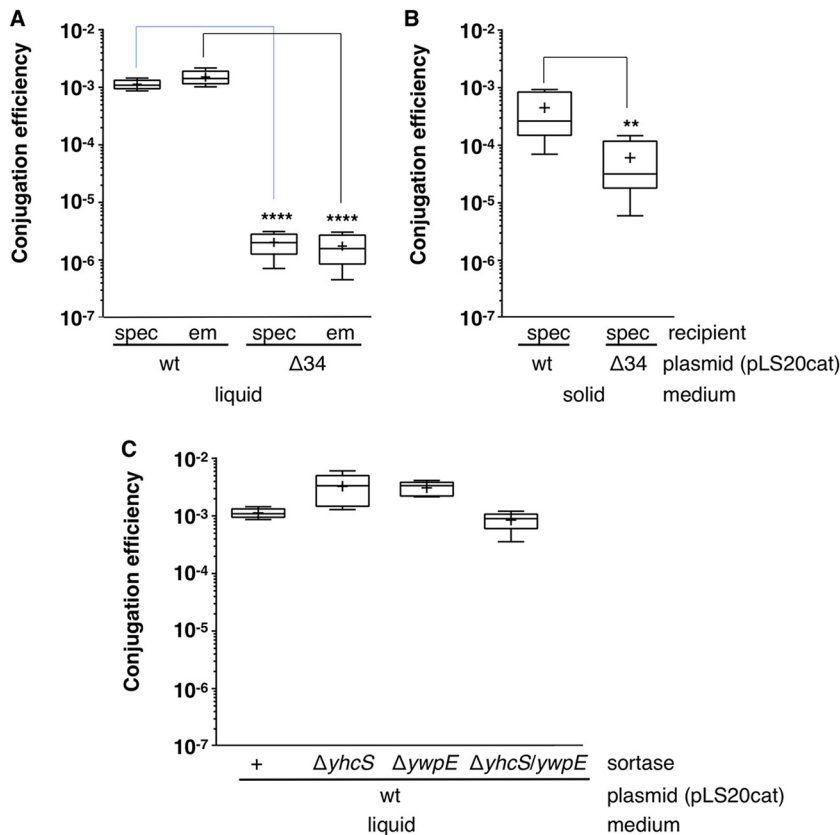


FIG 1 The absence of pLS20cat gene 34, but not the sortase genes *yhcS* and *ywpE*, affects conjugation efficiency more severely in liquid than in solid medium. (A and B) Conjugation efficiencies were determined for the wild-type (wt) plasmid pLS20cat (strain PKS11) and pLS20catΔ34 (strain CG164) in liquid (A) or solid (B) medium. To test possible differences due to the presence of different antibiotic markers, two strains with different resistance genes, PS110 (spectinomycin [spec] resistant) and PKS7 (erythromycin [em] resistant), were used as recipient strains in the liquid medium experiments. (C) Conjugation efficiencies in wild-type and sortase-deficient *yhcS* and/or *ywpE* donor strains. Conjugation efficiencies were calculated as the number of transconjugants per donor cell. Each experiment was repeated at least five times. Data are shown as box plot graphs. The box is determined by the 25th and 75th percentiles, and whiskers are determined by 5th and 95th percentiles; the line in the box indicates the median, and the “+” symbol indicates the mean for each sample data set. Analyses of variance (ANOVAs) show that the obtained differences in conjugation efficiencies between pLS20cat and pLS20catΔ34 are statistically significant with P values of $P < 0.0001$ (****) and $P < 0.01$ (**) for liquid and solid media, respectively. No significant differences were observed between different recipient strains used in liquid medium or in the sortase-negative strains ($P > 0.05$).

According to these analyses, protein p34 apparently contains a single N-terminal transmembrane-spanning helix (Fig. 3A, residues 9 to 31), which with 95% likelihood represented a signal peptide that would be cleaved between positions 35 and 36 (AEA-AT) by signal peptidase 1 (Fig. 3B). This shows that p34 has features of a protein that will be exported via the Sec-dependent secretion pathway.

Next, we performed a psi-blastp search of the NCBI nr database, using the deduced protein p34 sequence as a query (see Materials and Methods for details). After 16 rounds, this search resulted in the identification of 451 nonredundant hits showing significant similarity with the pLS20cat protein p34 (see Table S1 in the supplemental material).

Interestingly, 99.6% of the identified hits corresponded to proteins encoded by bacteria belonging to the phylum *Firmicutes*. A phylogenetic tree was calculated from these 451 hits, which revealed that the identified proteins could be divided into two clades (Fig. S1). Ninety-eight of the 194 hits of the first clade correspond to proteins encoded by bacteria belonging to the *Bacillus cereus* group, and most of these were

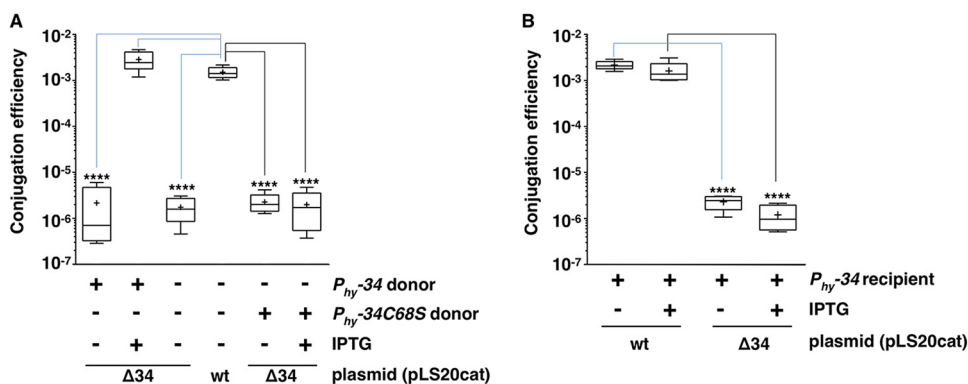


FIG 2 Effects of ectopic expression of the wild-type or the 34C68S mutated version of gene 34 in donor cells or recipient cells on pLS20cat and pLS20catΔ34 conjugation. Conjugation efficiencies were determined in liquid medium for pLS20cat and pLS20catΔ34, in the presence or absence of ectopic expression of the wild-type or the 34C68S mutated version of gene 34 in donor strains (A) or recipient strains (B). The following crosses were performed. For panel A (from left to right), CG159 × PKS7 without IPTG, CG159 × PKS7 with IPTG, CG164 × PS110, PKS11 × PS110, and CG203 × PKS7 without IPTG, and CG203 × PKS7 with IPTG; for panel B (from left to right), PKS11 × CG157 without IPTG, PKS11 × CG157 with IPTG, CG164 × CG157 without IPTG, and CG164 × CG157 with IPTG. Each experiment was repeated at least three times. ANOVAs show that the differences in conjugation efficiencies were statistically significant with P values of $P < 0.0001$ for matings of pLS20catΔ34 in the absence or presence of ectopically induced expression of gene 34 in the donor cells, but not in the recipient cells ($P > 0.05$). Although the differences were not statistically significant according to the ANOVA ($P = 0.0513$), a trend was observed that the conjugation levels obtained for the donor strain CG159 grown in the presence of IPTG were about 2-fold higher than those obtained for PKS11, the wild-type strain that harbored pLS20cat. See the legend to Fig. 1 for an explanation of the box plot graph symbols. Strains used: PS110, Spec^c; PKS7, Em^c; CG159, *amyE::P_{hyspank}-34*, pLS20catΔ34; CG164, pLS20catΔ34; PKS11, pLS20cat; CG157, *amyE::P_{hyspank}-34*; and CG203, *amyE::P_{hyspank}-34C68S*, pLS20catΔ34.

annotated as “fusion protein” (including the description “pXO2-28-29-30”). Protein p34 encoded by pLS20cat was assigned to clade 2, which included 257 proteins. Most of the clade 2 hits were annotated as hypothetical proteins. However, several were designated a (putative) function related to conjugation or adhesion, such as conjugal transfer protein (15 hits), isopeptide-forming-domain-containing fimbrial protein (12 hits), MucBP domain (19 hits), and thioester bond-forming surface protein (6 hits). These results suggest that many *Firmicutes* bacteria contain genes for proteins similar to that encoded by pLS20cat gene 34.

We next performed a blastp search against proteins encoded by plasmids present in the PLSDb database using stringent conditions (E value $< 1E-75$, coverage $> 75\%$), to see how many other plasmids encoded a protein similar to p34. This search resulted in the identification of 85 homologous proteins of which 22 had $> 75\%$ coverage, and 11 of these were encoded by unique plasmids (see Table 4 and Fig. S1). These 11 homologous genes were all present within putative conjugation operons (not shown), strongly indicating that these plasmids are conjugative and that the identified genes play similar roles in conjugation as gene 34 for pLS20cat.

In addition, we ran the protein p34 sequence against the HHpred server (24), applying standard parameters and selecting the PDB_mmCIF70 database. This revealed that the p34 region spanning residues 35 to 567 shared significant similarity with regions of 101 proteins, all encoded by G⁺ bacteria. Interestingly, all of them were annotated with one or more of the following keywords: surface protein, adhesin, collagen binding protein, LPXTG-anchored surface protein, thioester domain, pilin subunit, Ig-like fold, CnaA/CnaB folded domains, intramolecular amide bond, fimbrial, subunit, and/or integrin (Table S2). The hit with the highest level of homology and over the longest p34 region (residues 42 to 567, 99.8% probability, E value $1.1e-16$) corresponded to the *Bacillus anthracis*-encoded collagen adhesin protein named BaTIE (PDB 6FWV). This adhesin is a covalently cell wall-anchored protein with an N-terminal signal peptide, which is followed sequentially by a thioester domain (TED), three CnaB domains (see below), a

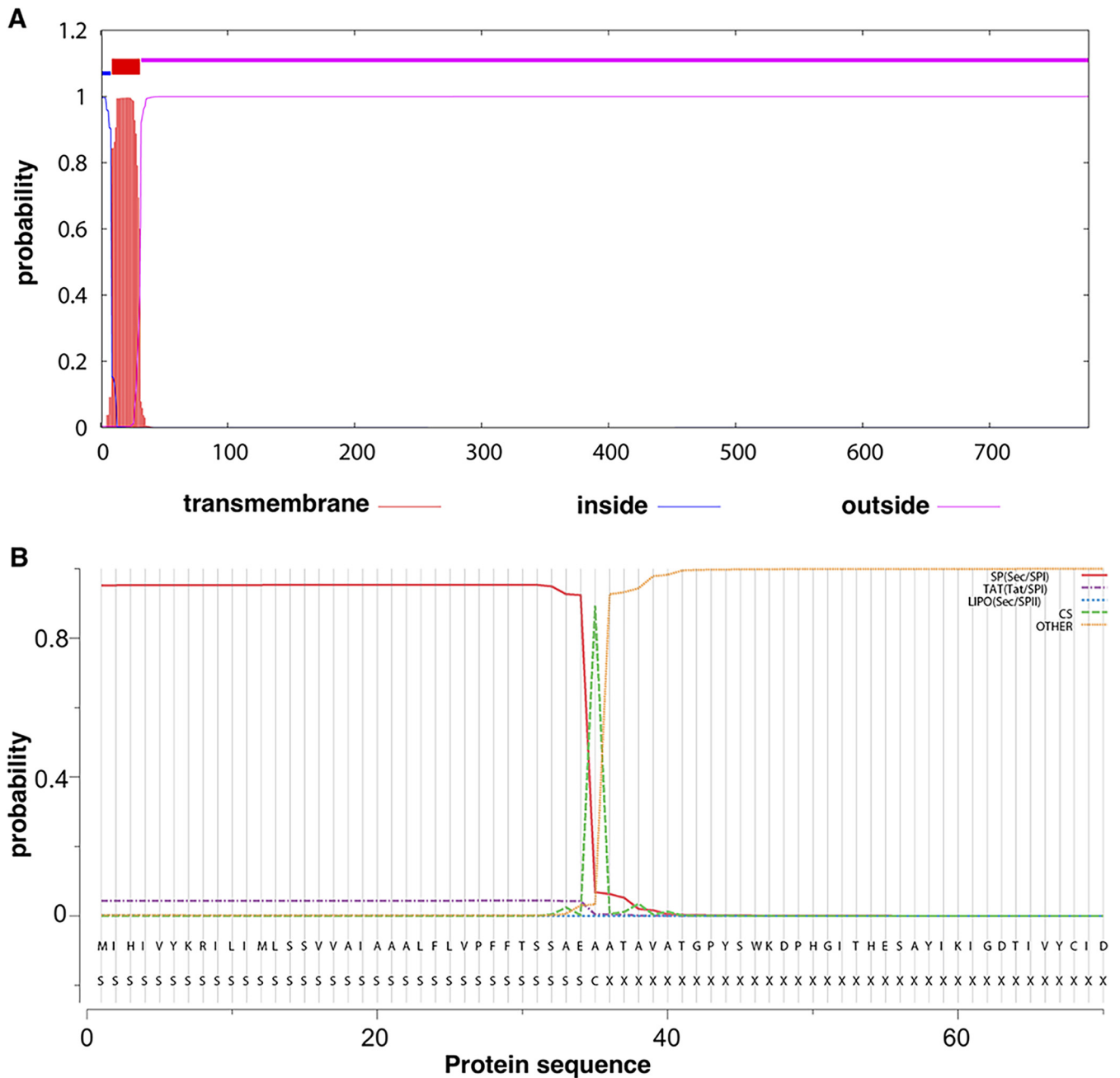


FIG 3 Localization and topology prediction of protein p34. (A) Prediction of transmembrane helices (red) and membrane topology of the pLS20cat-encoded protein p34 (of 778 residues) by the TMHMM 2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>) that uses a hidden Markov model (22, 47). (B) Protein p34 signal peptide prediction by the SignalP-5.0 server (<http://www.cbs.dtu.dk/services/SignalP/>) (23). The horizontal and vertical axes represent the p34 primary sequence and cleavage probability, respectively. The position of the predicted cleavage site (CS) is indicated with a dashed green line.

C-terminal LPXTG motif, and a transmembrane-spanning domain. The region of BaTIE predicted to share structural similarity with pLS20cat p34 corresponds to the TED and the three CnaB domains. These HHpred results, combined with the identification of a putative signal peptide and its importance in efficient conjugation in liquid medium, indicate that pLS20cat p34 is a cell wall-associated adhesin with a thioester and CnaA/CnaB domains. A large number of surface proteins from G⁺ bacteria containing a (putative) TED near their N terminus have now been identified. Often, these proteins are composed of multiple domains including TEDs, CnaA/B domains forming isopeptide bonds, and ester domains. Therefore, this family of proteins has been named TIE (thioester, isopeptide, ester) proteins (25). Based on this, we tentatively name pLS20cat gene 34 TIE_{pLS20}.

Evidence that TIE_{pLS20} (p34) contains a thioester domain followed by putative Cna domains. TEDs form a covalent thioester bond between a Cys and a Gln residue. The thioester-forming Cys residue is normally positioned within a four-residue conserved motif, [YFL]C ϕ π (ϕ and π corresponding to hydrophobic and hydrophilic residues, respectively) (25). The predicted TED of TIE_{pLS20} contains only one Cys residue (Cys68) that is embedded in the [YFL]C ϕ π motif “YCID” located within the predicted TED of TIE_{pLS20}. To study if Cys68 was important for TIE_{pLS20} function, the inducible, wild-type copy of *tie_{pLS20}* placed in the chromosome, which was used in the complementation experiment above, was replaced by a C68S mutant copy. The C68S mutation would prevent the formation of the Cys-Gln thioester bond by replacing the presumed reactive SH side chain of Cys with the hydroxyl group of Ser. The resulting strain, CG203 (*amyE::P_{hyspank}-tie_{pLS20}C68S, pLS20catΔ34*), was then used as donor in conjugation experiments, in parallel with the control donor strain CG159 containing a wild-type copy of *tie_{pLS20}* (*amyE::P_{hyspank}-tie_{pLS20}, pLS20catΔ34*). Efficient conjugation of pLS20catΔ34 was obtained upon ectopic expression of the wild-type but not the mutant version of *tie_{pLS20}* (Fig. 2A). These results demonstrate that residue Cys68 is essential for proper functioning of TIE_{pLS20}, probably by forming a thioester bond with a Gln residue, and hence that TIE_{pLS20} would possess a TED.

To determine which Gln residues would form the thioester bond with Cys68, we built a structural model of the presumed TED of TIE_{pLS20} based on the best hit obtained in the above-mentioned HHpred search: the *Bacillus anthracis* BaTIE protein which contains a class II type TED (26). The model was built with the program MODELLER (27) using as the template-query alignment the one provided by the HHpred search, comprising residues 44 to 567 of the primary TIE_{pLS20} sequence and BaTIE residues 12 to 516 of the 6FWV structure. The BaTIE structure used for building the model includes the TED (6FWV residues 1 to 258) followed by three CnaB domains (domains I, II, and III; 6FWV residues 259 to 343, 344 to 436, and 437 to 526, respectively). The model presented in Fig. 4 shows that Cys68 of TIE_{pLS20} would be located at the end of the slip-knot, and the Gln at bond-distance would be Gln256, indicating that Gln256 could form a thioester bond with Cys68. If Gln256 is functionally important, it was expected that this residue would be conserved. Inspection of the alignment of the 451 hits identified by the psi-blastp search revealed that p34 residue Gln256 is indeed highly conserved.

Most TEDs are located near the N terminus of the protein following a secretion signal. Generally, TED-containing adhesins have a large size and are composed of a variety of different domains and/or repeated sequences like isopeptide and ester domains, and fibronectin-binding or proline-rich repeats (25). In BaTIE, the thioester adhesion domain is followed by three structural domains adapting a characteristic sandwich-like structure that is similar to the so-called immunoglobulin domains (Ig domains or Ig-fold). Ig-folds are formed by two facing β -sheets, each being composed of antiparallel β -strands (Fig. 4A) (28). Many adhesins of G⁺ bacteria contain modified Ig-like domains, whose principal “stalk function” is to project the thioester adhesion domain away from the cell surface. The best-studied adhesins containing such folds are those similar to the collagen-binding Cna protein of *Staphylococcus aureus* (the Cna family) (29, 30), and therefore domains of adhesins containing such folds are also referred to as Cna domains. Interestingly, a characteristic feature of Ig-like folds, including Cna domains, is that, with the help of an adjacent catalytic Glu or Asp residue, they form intramolecular isopeptide bonds between a Lys and either an Asn or an Asp residue. These bonds are located at strategic positions within the protein and provide mechanical strength to the adhesin in order to resist shear forces, thereby ensuring firm adherence of the bacterial cell to its substrate. The isopeptide bonds are always in the hydrophobic interior of the domain, but depending on the strands that are joined, they are named CnaA or CnaB domains.

In BaTIE, the thioester adhesion domain is followed by three CnaB domains in which the isopeptide bonds are formed between Lys297 and Asn373 (domain I), Lys384 and Asp464 (domain II), and Lys475 and Asn555 (domain III). Residues Glu343, Glu443, and

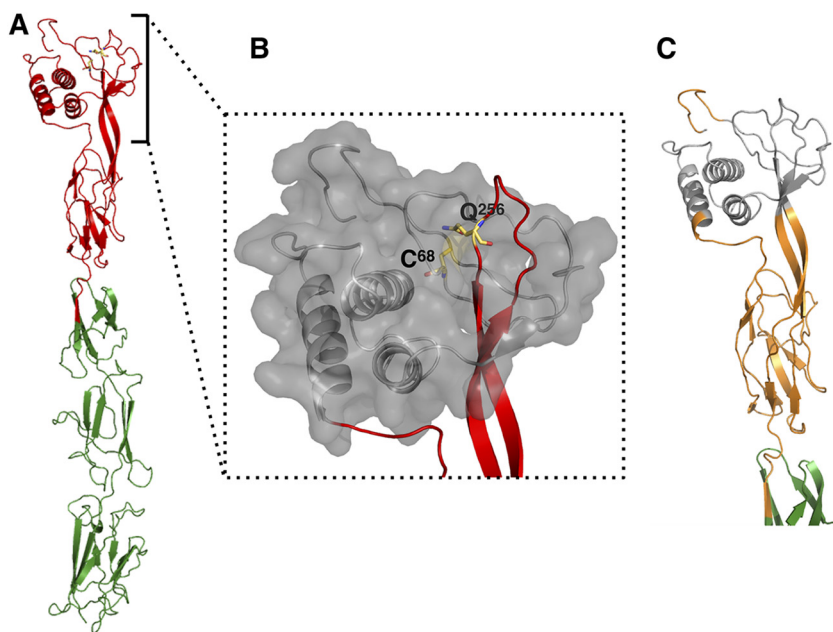


FIG 4 Computer-modeled structure of TIE_{pLS20} and its classification as a class II type TED. (A) Predicted structure built with the MODELLER program of TIE_{pLS20} (residues Tyr44 to Ala567) using as the template-query alignment BaTIE residues 12 to 516 (PDB ID 6FWV); the thioester domain (TED) is shown in red, the thioester-forming bonds Cys68 and Gln256 are shown as yellow sticks, and the Cna domains are shown in green. (B) Slipknot topology representation with its characteristic loop shown in red and the rest of the TED represented in gray with a transparent surface. (C) Structural elements that define TIE_{pLS20} as class II in the TED classification are shown in orange (based on the class II TED structural elements of BaTIE, PDB ID 6FWV) (26).

Glu524 are the putative catalytic residues for isopeptide bond formation in domains I, II, and III, respectively. According to HHpred, the TIE_{pLS20} region shared structural similarity with a continuous region of BaTIE encompassing the TED and the three subsequent CnaB domains. In the modeled structure of TIE_{pLS20}, the region following the thioester domain indeed has three Ig-like folds that are similar to the three CnaB domains in BaTIE. The corresponding domains in TIE_{pLS20} would consist of residues 286 to 349 (domain I), 350 to 438 (domain II), and 439 to 567 (domain III). However, it is presently hard to designate with high confidence the residues that form the isopeptide bonds.

In summary, TIE_{pLS20} shows a structural organization that is typically found in large multidomain TIE-type adhesins encoded by G⁺ bacteria. A schematic view of the structural organization of TIE_{pLS20} and the class II TED-containing TIE proteins BaTIE, SaTI, and EfmTIE86 is presented in Fig. 5.

TIE_{pLS20} is not coupled to the cell wall by a sortase. The structural similarity between TIE_{pLS20} and the cell wall-attached adhesins, such as BaTIE, raised the possibility that TIE_{pLS20} is also covalently attached to the cell wall by a sortase. Surprisingly, though, TIE_{pLS20} does not contain the typical cell wall sorting signal composed of the LPXTG motif at which the sortases act and a C-terminal transmembrane-spanning domain. To investigate this further, we tested whether sortases were required for the function of TIE_{pLS20}. *In silico* analysis showed that pLS20 did not contain a sortase gene. The bacterial genomes of most bacilli contain two genes, *yhcS* and *ywpE*, encoding probably functional but nonessential sortases (31, 32). Remarkably, the genome of *B. subtilis* strain 168 used in our studies contains a deletion affecting the first 81 to 82 codons of *ywpE* and its entire upstream gene. If a sortase is responsible for cell wall attachment of TIE_{pLS20}, the *yhcS* gene and/or the truncated *ywpE* gene would be expected to play a role in pLS20 conjugation. We therefore introduced pLS20cat into *yhcS* and *ywpE* single-mutant strains, and an *yhcS/ywpE* double mutant strain, and

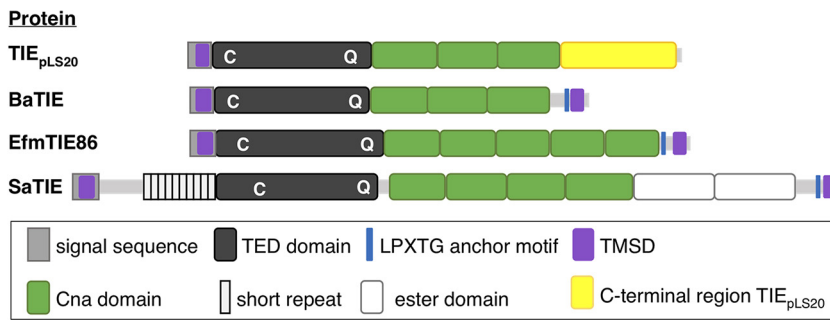


FIG 5 Structural organization of TIE_{pLS20} and TIE proteins encoded by the pathogens *B. anthracis*, *S. aureus*, and *E. faecium*, all containing a class II TED in their N-terminal region. The different domains are indicated with color-coded rectangles, and the proteins are aligned according to their thioester domains. Figures are drawn at scale. TIE_{pLS20}, BaTIE, EfmTIE86, and SaTIE are encoded by *B. subtilis* plasmid pLS20, *B. anthracis*, *E. faecium* VRE, and *S. aureus* VRS11b, respectively. The figure is adapted from Fig. 1A of the work of Miller et al. (26). TMSD, transmembrane-spanning domain.

used these as donors in conjugation experiments. As shown in Fig. 1C, high levels of conjugation, similar to those observed for the wild-type donor strain PKS11, were observed. These results show that neither the host-encoded sortase YhcS nor the truncated YwpE proteins are required for linking TIE_{pLS20} to the cell wall.

DISCUSSION

Mating pair formation is a crucial initial step in the conjugation process and can thus be a target for combating conjugation-mediated spread of antibiotic resistance and virulence genes. So far, little is known about mating pair formation in *G*⁺ bacteria. The only mating pair system in *G*⁺ bacteria that has been studied in considerable depth is encoded by the conjugative enterococcal plasmid pCF10 (14). The pCF10-encoded PrgB surface protein is involved in forming mating pair aggregates that are important for efficient conjugation in liquid cultures but not on solid surfaces. It is also a virulence factor promoting attachment and biofilm development of *E. faecalis* cells on biotic and abiotic surfaces. The PrgB protein is exported via an N-terminal secretion signal and is anchored to the cell wall by its C-proximally located LPXTG motif. One adhesin and two Arg-Gly-Asp (RGD) domains are located in between the signal peptide and the LPXTG motif. The RGD-containing domains are implicated in binding host cell integrins. Cell aggregation and efficient conjugation require, besides PrgB, also extracellular DNA (eDNA) (33). The structure of the PrgB adhesin domain that is responsible for eDNA binding and compaction shares similarity to the lectin-like fold in general and, particularly, to the multimodal *Streptococcus* sp. AgII/II, SspB, and SpaP adhesins (15, 34). It has been proposed that PrgB-mediated compaction of eDNA may serve an analogous function as the retractile conjugative F-pili. Since the *G*⁺ cell wall component lipoteichoic acid (LTA) competes with eDNA for binding with the adhesion domain, it is envisioned that PrgB first favors cell-cell contact through eDNA compaction and then stabilizes these contacts through LTA binding. Finally, proper functioning of PrgB probably requires also the surface protein PrgA in a yet-unknown manner. Our results show that, in the case of pLS20, mating pair formation is also facilitated by an adhesin encoded by gene 34 of pLS20, which we have named *tie*_{pLS20}. Like PrgB, TIE_{pLS20} is important for efficient conjugation, particularly in liquid medium. However, TIE_{pLS20} is structurally very different from PrgB. TIE_{pLS20} has a typical *G*⁺ adhesin domain architecture with an N-terminal secretion signal, followed by a class II type thioester domain and three structural domains whose functions are probably to direct the adhesin domain away from the donor surface and to provide strength and stability to the large elongated adhesin molecule.

To exert their function, after export from the cytoplasm, surface proteins must be retained on the cell wall. In *G*⁺ bacteria, there are three known mechanisms that link

surface proteins to the cell wall (for review, see references 35 to 37). First, lipoproteins contain a lipid moiety linked to their N-terminal Cys residue. TIE_{pLS20} does not contain a lipobox, making it highly unlikely that it will become attached to the membrane by N-terminal lipidation. Another reason making this scenario unlikely is that this would position the TED adhesion domain either (partially) in the cell wall or at the cell surface with the Cna domains pointing away from the cell surface. A second way of linking surface proteins to the cell wall is through modules that mediate multiple noncovalent hydrophobic and/or charge-based interactions with cell wall components. Examples of such modules are the GW module, LysM motif, or surface-layer homology domain. TIE_{pLS20} does not contain a module or region showing similarity with any such domains. Finally, a third way by which surface proteins can become attached to the cell wall is through sortase-mediated covalent attachment to the peptidoglycan. These surface proteins contain, besides the N-terminal secretion signal, a C-terminal sorting signal consisting of a conserved LPXTG motif followed by a C-terminal transmembrane domain and a positively charged tail. Sortases act on the LPXTG motif and either link proteins together to form pili or attach large multidomain proteins like TIE proteins to cross-bridge peptides of the cell wall, which are then incorporated as precursors into the cell wall by penicillin-binding proteins. Most surface proteins, including adhesins belonging to the family of “microbial surface components recognizing adhesive matrix molecules” (MSCRAMM) or pilin structures, are covalently anchored to the cell wall in a sortase-dependent way (30, 38, 39). Sortases are also known to attach the three adhesins BaTIE, SaTIE, and EfmTIE86 containing the class II type TED on the cell surface. However, our results indicate that TIE_{pLS20} does not become anchored to the cell wall in a sortase-dependent manner: pLS20 does not carry a sortase gene, and we showed that the conjugation efficiency was not significantly affected using donor strains lacking one or both of the chromosomally located sortase genes. Moreover, TIE_{pLS20} lacks a typical sorting signal. Therefore, the absence of a lipobox, sorting signal, and any known modules that would allow noncovalent interactions with cell wall components raises the intriguing question of how TIE_{pLS20} attaches to the cell wall. The conserved structural organization of G⁺ adhesins and that of TIE_{pLS20} suggest that TIE_{pLS20} remains attached to the cell wall through its C-terminal region, which will direct the TED adhesion module away from the cell surface. According to the RaptorX protein modeling server (40), part of the approximately 250-residue C-terminal region of TIE_{pLS20} shares similarity with a structural domain of unknown function named “toast-rack” (pfam17115) that is present on a putative adhesin encoded by *Clostridium sporogenes* ATCC 15579 (PDB 4QRK). Although at present we do not know how TIE_{pLS20} becomes attached to the cell wall, based on all the arguments outlined above, it seems plausible that the C-terminal region of TIE_{pLS20} plays a role in this. Future subcellular localization studies are required to verify this hypothesis. Likewise, it will be interesting to investigate how TIE_{pLS20} is targeted to the *B. subtilis* cell envelope and how exactly this protein facilitates direct contact between the donor and recipient cells for conjugal DNA transfer.

We have also shown that TIE_{pLS20} contains a class II type TED, which is probably responsible for the adhesive properties of the protein. The constructed model indicated that the TIE_{pLS20} residues Cys68 and Gln256 form a thioester bond, which was supported by the fact that the *tie_{pLS20}C68S* mutant is not functional in conjugation. Surface-located adhesins and their crucial roles in attachment of commensal and pathogenic bacteria to host surfaces as a prerequisite for colonization and infection have been known for a long time (41). In many of these cases, adhesion of bacteria to a host is based on noncovalent interactions, involving extensive intermolecular regions. The possibility of covalent interaction of a bacterium with a host through a reactive thioester bond was discovered only relatively recently. In 2010, the Cpa pilus tip adhesin from the G⁺ human pathogen *Streptococcus pyogenes* was shown to contain a thioester bond required for efficient host cell interaction (42). In 2015, Walden et al. showed that many G⁺ adhesins contain a (putative) TED and hence that covalent attachment to the host is more common than previously assumed (25). They also revealed structures of class I TEDs and demonstrated that the streptococcal adhesin SfbI reacts with one specific fibrinogen lysine residue in a thioester-dependent

TABLE 1 Strains used

Bacterium	Strain	Genotype	Reference and/or source
<i>E. coli</i>	XL1-Blue	<i>endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44</i> <i>F' [::Tn10 proAB+ lacIq Δ(lacZ)M15]</i> <i>hsdR17(rK- mK+)</i>	56
	JM101	<i>F' traD36 proA+B+ lacIq Δ(lacZ)M15/Δ(lac-proAB)</i> <i>glnV thi</i>	57
<i>B. subtilis</i>	168 (1A700)	<i>trpC2</i>	BGSC ^a
	PS110	<i>trpC2, amyE::P_{spank}-Δ (spec)</i>	19
	PKS7	<i>trpC2, thrC::Em</i>	19
	PKS11	<i>trpC2, pLS20cat</i>	58
	PKS91	<i>trpC2, pLS20spec</i>	21
	CG164	<i>trpC2, pLS20catΔ34</i>	This work
	CG157	<i>trpC2, amyE::P_{hyspank}-34/tie_{pLS20} (spec)</i>	This work
	CG202	<i>trpC2, amyE::P_{hyspank}-34/tie_{pLS20}C68S (spec)</i>	This work
	CG159	<i>trpC2, amyE::P_{hyspank}-34/tie_{pLS20} (spec), pLS20catΔ34 (cat)</i>	This work
	CG203	<i>trpC2, amyE::P_{hyspank}-34/tie_{pLS20}C68S (spec), pLS20catΔ34 (cat)</i>	This work
	BKE09200	<i>trpC2, yhcS::Em</i>	BGSC, 59
	BKK09200	<i>trpC2, yhc::Km</i>	BGSC, 59
	BKE36340	<i>trpC2, ywpE::Em</i>	BGSC, 59
	BKK36340	<i>trpC2, ywpE::Km</i>	BGSC, 59
	CG265	<i>trpC2, yhcS::Em, pLS20cat</i>	This work
	CG266	<i>trpC2, yhcS::Km, pLS20cat</i>	This work
	CG267	<i>trpC2, ywpE::Em, pLS20cat</i>	This work
	CG268	<i>trpC2, ywpE::Km, pLS20cat</i>	This work
	CG269	<i>trpC2, ywpE::Km, yhcS::Em</i>	This work
	CG271	<i>trpC2, ywpE::Km, yhcS::Em, pLS20cat</i>	This work

^aBGSC, Bacillus Genetic Stock Centre.

mechanism, resulting in a very stable intermolecular amide bond (25). Determination of class II TED structures revealed that they contain, compared to class I TEDs, an additional β -sandwich domain that forms a slipknot structure with the conserved TED fold. The structures of the class II TEDs have been reported for three pathogens, *B. anthracis* and vancomycin-resistant strains of *S. aureus* and *Enterococcus faecium* (26). TEDs thus appear to be common adhesin molecules permitting covalent attachment of many pathogens to host cells. However, this is the first time that a TED-containing adhesin has been shown to play a role in conjugation.

Database searches revealed that several other conjugative plasmids of G⁺ bacteria also contain a class II type TED-containing adhesin, suggesting that mating pair formation is also mediated by the TED domain in these cases. In fact, it is probable that mating pair formation in G⁺ bacteria is mediated by adhesins in general. In this respect, it

TABLE 2 Plasmids used

Plasmid name	Description	Reference
pLS20cat	Native plasmid pLS20 labeled with Cm resistance gene in unique Sall site	Laboratory stock
pLS20catΔ34	Derivative of pLS20cat containing large internal deletion of gene <i>34/tie_{pLS20}</i>	This work
pMiniMAD2	Plasmid used for markerless deletions	Gift of Daniel Kearns
pCGdelta34	pMiniMAD2 derivative to create in-frame markerless partial deletion (from codon 285 to 605) of pLS20cat gene <i>34/tie_{pLS20}</i>	This work
pDR110	<i>B. subtilis amyE</i> integration vector containing IPTG-inducible P _{spank} promoter	Gift of David Rudner
pDR111	<i>B. subtilis amyE</i> integration vector containing IPTG-inducible P _{hyspank} promoter	Gift of David Rudner
pCG197	pDR111 derivative containing a 729-bp fragment coding for the N-terminal region of gene <i>34/tie_{pLS20}</i> harboring the C68S mutation behind the P _{hyspank} promoter	This work
pCG205	pDR111 derivative containing the complete pLS20cat gene <i>34/tie_{pLS20}</i> harboring C68S mutation behind the P _{hyspank} promoter	This work

TABLE 3 Oligonucleotides used

Name	Sequence (5'–3') ^a	Purpose
oEST15	tttt gtcgac GACAATAAGAAAGGAGGTGATCGAAGTGATTC	Forward oligo ^b to amplify gene <i>34/tie_{pLS20}</i> and to be cloned in pDR110/11, includes Sall site
oEST16	tttt gtcgac CTCCAAACAGTTGAAAAGTTATTTCTCAACC	Reverse oligo to amplify gene <i>34/tie_{pLS20}</i> and to be cloned in pDR110/11, includes SphI site
pDR111_U_sec	TGACTTTATCTACAAGGTGTGGC	Oligo used for sequencing, and colony PCR of derivatives of pDR110/11
pDR111_L_sec	TTAAATGCAACCGTTTTTCGGAAGG	Oligo used for sequencing, check colony PCR of derivatives of pDR110/11
LS20_8	AGTTATGACAGAAGTAACCCCAACA	Internal oligo of gene <i>34/tie_{pLS20}</i> used for sequencing, and colony PCR
LS20_4Back	CTTCCATGAATACGGACCTGTGGC	Internal oligo of gene <i>34/tie_{pLS20}</i> used for sequencing, and colony PCR
oCG49	CCCCATCGGGATTGTCTTTG	Oligo used for sequencing and colony PCR of derivatives of pMiniMAD2
oCG50	aaa agtcgac GGTGACGGTAATGAAATTGGA	Forward oligo to amplify <i>34/tie_{pLS20}</i> "UP" region in combination with oligo oCG51. It also contains an NheI restriction site extension used for cloning of the PCR fragment in pMiniMAD2
oCG51	CTCTTTACTTTCTTAAAGACCGTGCCTTCGAATTTGATATCATCATCGTC	Reverse oligo to amplify <i>34/tie_{pLS20}</i> "UP" region in combination with oligo oCG50. Contains a 5' extension used in subsequent overlapping PCR to fuse the <i>34/tie_{pLS20}</i> "UP" region with the "DOWN" region
oCG52	GACGATGATGATATCAAATTCGATTTAAAGTACCTGAAG	Forward oligo to amplify <i>34/tie_{pLS20}</i> "DOWN" region in combination with oligo oCG53. Contains a 5' extension used in subsequent overlapping PCR to fuse the <i>34/tie_{pLS20}</i> "DOWN" region with the "UP" region
oCG53	tttt ggatcc GAGTCTCGTTATTTTTGTGATGGC	Reverse oligo to amplify <i>34/tie_{pLS20}</i> "DOWN" region in combination with oligo oCG52. It also contains a BamHI restriction site extension used for cloning of the PCR fragment in pMiniMAD2
oCG54	GTCTACTTAGCTCTTTTTTACAATAC	Oligo used for sequence analysis and colony PCR of derivatives of pMiniMAD2
oCG124	GCTTATCTGGATCAATACTGTAACAATGGTGTCTCTCTATTTTGATGT AAGCTGATTCATGAGTAATGC	Reverse oligo to generate "UP" fragment that contains the mutation in gene <i>34/tie_{pLS20}</i> to change codon Cys68 into Ser68 (indicated in italic). Used in combination with oEST15
oCG125	CCATTGTTTACAGTATTGATCCAGATAAGCCTGCACCTTATGGCGTGC ATTCGTATAAAACCCCGAAGCGT	Forward oligo to generate "DOWN" fragment that contains the mutation in gene <i>34/tie_{pLS20}</i> where codon Cys68 is substituted by Ser68 (indicated in italic). Used in combination with oCG126
oCG126	aaa agcatgct taCGTAGTCAGATCCATAATCTAAAGGTGC	Reverse oligo to amplify the "DOWN" region of gene <i>34/tie_{pLS20}</i> . It contains a 5' extension with SphI restriction site; used in combination with oCG125
oCG141	GGCAGCTCTGTTCCGGTGATCAGCAAAAATCAGAAACCATTAAGC	Forward oligo to amplify gene <i>34/tie_{pLS20}</i> starting from codon 167. Used in combination with oEST16

^aCapital letters, pLS20 sequences; bold, restriction enzyme sites; underlining, stop codons (also underlined); italic, overlapping sequences.

^boligo, oligonucleotide.

is interesting that antiadhesion therapies have been studied for more than 2 decades as a way to combat bacterial infections that are resistant to antibiotics. These therapies include the use of receptor and adhesin analogues, dietary constituents, and adhesin-based vaccines (43, 44). It is possible that similar strategies can be applied to curtail adhesin-based mating pair formation, thereby impeding conjugation-mediated spread of antibiotic resistance and virulence genes.

MATERIALS AND METHODS

Bacterial strains, plasmids, media, and oligonucleotides. *B. subtilis* and *E. coli* strains were grown and selected in lysogeny broth (LB), or on 1.5% LB agar plates. When appropriate, media were supplemented with the following antibiotics: ampicillin (100 μg/ml), spectinomycin (100 μg/ml), chloramphenicol (5 μg/ml), erythromycin (1 and 150 μg/ml for *B. subtilis* and *E. coli*, respectively), and kanamycin (10 and 30 μg/ml for *B. subtilis* and *E. coli*, respectively). For induction of the P_{spank} and P_{hyspank} promoters and the P_{xyI} promoter, media were supplemented with 1 mM isopropyl-β-D-1-thiogalactopyranoside

TABLE 4 Conjugative plasmids of G⁺ bacteria encoding a pLS20cat p34 homologue

Bacterial strain	Plasmid name	Plasmid identifier	Protein identifier	% identity	Length alignment (residues)	MM ^a	Gaps	q start ^b	q end ^c	S start ^d	S end ^e	E value	Bit score ^f
<i>B. subtilis</i> natto	pLS20	NC_015148.1	YP_004243498.1	99.9	753	1	0	26	778	1	753	0	1139
<i>B. subtilis</i> KH2	NG ^g	NZ_CP018185.1	WP_033881018.1	99.9	777	1	0	2	778	1	777	0	1179
<i>B. subtilis</i> SRCM100333	pBS333	NZ_CP021893.1	WP_088272629.1	100	767	0	0	12	778	1	767	0	1160
<i>Bacillus velezensis</i> DSYZ	pDSYZ	NZ_CP030151.1	WP_082188812.1	77.7	767	171	0	12	778	1	767	0	1146
<i>Listeria grayi</i>	pLGUG1	NC_014496.1	YP_003877896.1	51.4	775	370	3	4	778	2	769	0	1145
<i>Listeria monocytogenes</i> R479a	pLMR479a	NZ_HG813248.1	WP_025370673.1	51.0	774	372	3	4	777	2	768	0	1137
<i>Listeria monocytogenes</i> FDA0006907	pCFSAN021445	NZ_CP022021.1	WP_003725232.1	51.9	774	363	6	5	778	3	767	0	1102
<i>Listeria monocytogenes</i> HPB5415	pLM5578	NZ_CP019166.1	WP_012952147.1	51.7	770	362	6	11	777	7	769	0	1091
<i>Bacillus velezensis</i> GH1-13	NG ^g	NZ_CP019039.1	WP_077721679.1	58.7	762	308	6	20	778	3	760	0	1063
<i>Bacillus safensis</i> U14-5	NG ^g	NZ_CP015608.1	WP_075623813.1	52.3	767	359	6	15	778	8	770	0	1035
<i>Terribacillus goriensis</i> MP602	pT1	NZ_CP008877.1	WP_041592320.1	35.1	764	473	12	30	777	111	867	0	885

^aMM, mismatches.

^bStart alignment query sequence.

^cEnd alignment query sequence.

^dStart sequence hit.

^eEnd sequence.

^fBit score.

^gNot given. Note that this search identified the p34 sequence of pLS20cat as deposited in the NCBI database, and hence, it is 25 residues shorter (see first paragraph of Results).

(IPTG) or 1% xylose, respectively. Antibiotics and chemicals were purchased from Sigma-Aldrich, oligonucleotides were from Isogen (Life Science, The Netherlands), and restriction enzymes and T4 DNA ligase were from New England Biolabs (NEB). All *B. subtilis* strains used were isogenic with *B. subtilis* strain 168 (Bacillus Genetic Stock Centre [Table 1]). Plasmids and oligonucleotides are listed in Tables 2 and 3, respectively.

Transformation. *E. coli* cells were transformed by standard procedures (45). Generation of competent *B. subtilis* cells and natural transformations were done as previously described (46).

Conjugation assays. Conjugation in liquid medium was carried out as described previously (19). Conjugation in solid medium was performed similarly but with the following changes. Samples of late exponentially growing donor and recipient cultures (optical density at 600 nm [OD₆₀₀] between 0.8 and 1.0) were mixed at 1:1 stoichiometry and spread on nonselective plates. Part of the mixture was plated on selective plates, to verify the 1:1 stoichiometry of donor and recipient cells. Cells grown overnight on a nonselective plate were harvested, and dilutions were spread on selective plates to select for donors and transconjugants.

Construction of plasmids and strains. Details on the construction of plasmids and strains are given in Text S1 in the supplemental material. Table 4 shows conjugative plasmids of G⁺ bacteria encoding a pLS20cat p34 homologue.

In silico analyses. (i) Identification of membrane, secreted, and surface proteins. Deduced pLS20cat protein sequences were screened for the presence of transmembrane-spanning domains and their transmembrane topology using the TMHMM server version 2.0 of the Centre for Biological Sequence Analyses at the Danish Technical University (DTU; www.cbs.dtu.dk/services/TMHMM) (22, 47). The presence of potential signal peptidase 1 cleavage sites was predicted using the SignalP-5.0 server (<http://www.cbs.dtu.dk/services/SignalP/>) with default settings for proteins of G⁺ bacteria.

(ii) Identification of genes encoding proteins with significant similarity to pLS20cat protein p34. The primary sequence of the pLS20cat-encoded protein p34 was used as a query sequence to execute psi-blastp searches (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) against the NCBI nr protein database (version 2.10.0+, February 2020) (48–50). Iterative rounds of psiBlast searches were done until no new sequences with an E value of <1E–75 and a coverage of >75% of the entire length of the query sequence were incorporated into the psiBlast profile, which occurred after 16 rounds. All sequences with an E value of <1E–20 and a coverage of >75% were retrieved for subsequent analysis.

(iii) Generation of a phylogenetic tree. First, protein sequences of all 451 proteins with similarity to p34 were aligned using the program “decipher” with default settings (51). Next, an average linked tree was built using blosum62 similarity score as distance with the program Jalview (52, 53). The resulting phylogenetic tree was plotted with Dendroscope (54).

(iv) Identifying protein p34 homologues encoded by plasmids. We generated a database containing all protein sequences encoded by plasmids present in the PLSDDB, a plasmid database (55). Next, the p34 protein sequence was used as query to perform a blastp search against this protein database.

(v) Structure prediction and modeling. Sequences of interest were submitted for online predictions to the following servers: HHpred (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) (24), MODELLER

(<https://salilab.org/modeller/>) (27), and RaptorX (<http://raptorx.uchicago.edu/>) (40). Analyses were performed using standard parameters.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.02 MB.

FIG S1, PDF file, 0.5 MB.

TABLE S1, XLSX file, 0.02 MB.

TABLE S2, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

This work was supported by Ministry of Economy and Competitiveness of the Spanish Government PID2019_108778GB_C21 (AEI/FEDER, EU) to W.J.J.M., a FEMS Research and Training Grant (FEMS-GO-2018-222) to C.G.C., and a Wellcome Investigator grant [209500] to Jeff Errington that supported L.J.W. R.A.S. acknowledges support through a National Council of Science and Technology scholarship (CONACyT grant 428302). This research was also supported by institutional grants from the Fundación Ramón Areces and Banco de Santander to the Centro de Biología Molecular Severo Ochoa.

We thank Daniel Zeigler of the Bacillus Genetic Stock Centre for sending us the sortase mutants and Ana Martín María and David Romera García for their help with conjugation experiments and bioinformatics, respectively.

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304. <https://doi.org/10.1038/35012500>.
- Frost LS, Lepplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 3:722–732. <https://doi.org/10.1038/nrmicro1235>.
- Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 3:711–721. <https://doi.org/10.1038/nrmicro1234>.
- Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat Rev Genet* 16:472–482. <https://doi.org/10.1038/nrg3962>.
- Mazel D, Davies J. 1999. Antibiotic resistance in microbes. *Cell Mol Life Sci* 56:742–754. <https://doi.org/10.1007/s000180050021>.
- World Health Organization. 2017. WHO's first global report on antibiotic resistance reveals serious, worldwide threat to public health. World Health Organization, Geneva, Switzerland. <http://www.who.int/mediacentre/news/releases/2014/amr-report/en/>.
- Graf FE, Palm M, Warringer J, Farewell A. 2019. Inhibiting conjugation as a tool in the fight against antibiotic resistance. *Drug Dev Res* 80:19–23. <https://doi.org/10.1002/ddr.21457>.
- Lerminiaux NA, Cameron ADS. 2019. Horizontal transfer of antibiotic resistance genes in clinical environments. *Can J Microbiol* 65:34–44. <https://doi.org/10.1139/cjm-2018-0275>.
- Li YG, Christie PJ. 2018. The Agrobacterium VirB/VirD4 T4SS: mechanism and architecture defined through in vivo mutagenesis and chimeric systems. *Curr Top Microbiol Immunol* 418:233–260. https://doi.org/10.1007/82_2018_94.
- Thompson MA, Onyeziri MC, Fuqua C. 2018. Function and regulation of Agrobacterium tumefaciens cell surface structures that promote attachment. *Curr Top Microbiol Immunol* 418:143–184. https://doi.org/10.1007/82_2018_96.
- Arutyunov D, Frost LS. 2013. F conjugation: back to the beginning. *Plasmid* 70:18–32. <https://doi.org/10.1016/j.plasmid.2013.03.010>.
- Li YG, Hu B, Christie PJ. 2019. Biological and structural diversity of type IV secretion systems. *Microbiol Spectr* 7(2). <https://doi.org/10.1128/microbiolspec.PSIB-0012-2018>.
- Hu B, Khara P, Christie PJ. 2019. Structural bases for F plasmid conjugation and F pilus biogenesis in Escherichia coli. *Proc Natl Acad Sci U S A* 116:14222–14227. <https://doi.org/10.1073/pnas.1904428116>.
- Dunny GM. 2013. Enterococcal sex pheromones: signaling, social behavior, and evolution. *Annu Rev Genet* 47:457–482. <https://doi.org/10.1146/annurev-genet-111212-133449>.
- Schmitt A, Jiang K, Camacho MI, Jonna VR, Hofer A, Westerlund F, Christie PJ, Berntsson RP. 2018. PrgB promotes aggregation, biofilm formation, and conjugation through DNA binding and compaction. *Mol Microbiol* 109:291–305. <https://doi.org/10.1111/mmi.13980>.
- Itaya M, Sakaya N, Matsunaga S, Fujita K, Kaneko S. 2006. Conjugational transfer kinetics of pLS20 between Bacillus subtilis in liquid medium. *Bio-sci Biotechnol Biochem* 70:740–742. <https://doi.org/10.1271/bbb.70.740>.
- Koehler TM, Thorne CB. 1987. Bacillus subtilis (natto) plasmid pLS20 mediates interspecies plasmid transfer. *J Bacteriol* 169:5271–5278. <https://doi.org/10.1128/jb.169.11.5271-5278.1987>.
- Meijer WJJ, Wisman GBA, Terpstra P, Thorsted PB, Thomas CM, Holsappel S, Venema G, Bron S. 1998. Rolling-circle plasmids from Bacillus subtilis: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from Gram-positive bacteria. *FEMS Microbiol Rev* 21:337–368. <https://doi.org/10.1111/j.1574-6976.1998.tb00357.x>.
- Singh PK, Ramachandran G, Ramos-Ruiz R, Peiro-Pastor R, Abia D, Wu LJ, Meijer WJ. 2013. Mobility of the native Bacillus subtilis conjugative plasmid pLS20 is regulated by intercellular signaling. *PLoS Genet* 9:e1003892. <https://doi.org/10.1371/journal.pgen.1003892>.
- Ramachandran G, Singh PK, Luque-Ortega JR, Yuste L, Alfonso C, Rojo F, Wu LJ, Meijer WJ. 2014. A complex genetic switch involving overlapping divergent promoters and DNA looping regulates expression of conjugation genes of a gram-positive plasmid. *PLoS Genet* 10:e1004733. <https://doi.org/10.1371/journal.pgen.1004733>.
- Gago-Córdoba C, Val-Calvo J, Miguel-Arribas A, Serrano E, Singh PK, Abia D, Wu LJ, Meijer WJJ. 2019. Surface exclusion revisited: function related to differential expression of the surface exclusion system of Bacillus subtilis plasmid pLS20. *Front Microbiol* 10:1502. <https://doi.org/10.3389/fmicb.2019.01502>.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to

- complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
23. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
 24. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430:2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>.
 25. Walden M, Edwards JM, Dziewulska AM, Bergmann R, Saalbach G, Kan SY, Millro OK, Weckener M, Jackson RJ, Shirran SL, Botting CH, Florence GJ, Rohde M, Banfield MJ, Schwarz-Linek U. 2015. An internal thioester in a pathogen surface protein mediates covalent host binding. *Elife* 4:e06638. <https://doi.org/10.7554/eLife.06638>.
 26. Miller OK, Banfield MJ, Schwarz-Linek U. 2018. A new structural class of bacterial thioester domains reveals a slipknot topology. *Protein Sci* 27:1651–1660. <https://doi.org/10.1002/pro.3478>.
 27. Webb B, Sali A. 2016. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 86:2.9.1–2.9.37. <https://doi.org/10.1002/cpps.20>.
 28. Berisio R, Ciccarelli L, Squeglia F, De Simone A, Vitagliano L. 2012. Structural and dynamic properties of incomplete immunoglobulin-like fold domains. *Protein Pept Lett* 19:1045–1053. <https://doi.org/10.2174/092986612802762732>.
 29. Foster TJ. 2019. The MSCRAMM family of cell-wall-anchored surface proteins of gram-positive cocci. *Trends Microbiol* 27:927–941. <https://doi.org/10.1016/j.tim.2019.06.007>.
 30. Baker EN, Squire CJ, Young PG. 2015. Self-generated covalent cross-links in the cell-surface adhesins of Gram-positive bacteria. *Biochem Soc Trans* 43:787–794. <https://doi.org/10.1042/BST20150066>.
 31. Fasehee H, Westers H, Bolhuis A, Antelmann H, Hecker M, Quax WJ, Mirolohi AF, van Dijk JM, Ahmadian G. 2011. Functional analysis of the sortase YhcS in *Bacillus subtilis*. *Proteomics* 11:3905–3913. <https://doi.org/10.1002/pmic.201100174>.
 32. Liew PX, Wang CL, Wong SL. 2012. Functional characterization and localization of a *Bacillus subtilis* sortase and its substrate and use of this sortase system to covalently anchor a heterologous protein to the *B. subtilis* cell wall for surface display. *J Bacteriol* 194:161–175. <https://doi.org/10.1128/JB.05711-11>.
 33. Bhatti M, Cruz MR, Frank KL, Gomez JA, Andrade F, Garsin DA, Dunny GM, Kaplan HB, Christie PJ. 2015. Enterococcus faecalis pCF10-encoded surface proteins PrgA, PrgB (aggregation substance) and PrgC contribute to plasmid transfer, biofilm formation and virulence. *Mol Microbiol* 95:660–677. <https://doi.org/10.1111/mmi.12893>.
 34. Bhatti M, Camacho MI, Gonzalez-Rivera C, Frank KL, Dale JL, Manias DA, Dunny GM, Christie PJ. 2017. PrgU: a suppressor of sex pheromone toxicity in *Enterococcus faecalis*. *Mol Microbiol* 103:398–412. <https://doi.org/10.1111/mmi.13563>.
 35. Dramsi S, Magnet S, Davison S, Arthur M. 2008. Covalent attachment of proteins to peptidoglycan. *FEMS Microbiol Rev* 32:307–320. <https://doi.org/10.1111/j.1574-6976.2008.00102.x>.
 36. Buist G, Steen A, Kok J, Kuipers OP. 2008. LysM, a widely distributed protein motif for binding to (peptidoglycans). *Mol Microbiol* 68:838–847. <https://doi.org/10.1111/j.1365-2958.2008.06211.x>.
 37. Nguyen MT, Matsuo M, Niemann S, Herrmann M, Gotz F. 2020. Lipoproteins in Gram-positive bacteria: abundance, function, fitness. *Front Microbiol* 11:582582. <https://doi.org/10.3389/fmicb.2020.582582>.
 38. Hendrickx AP, Budzik JM, Oh SY, Schneewind O. 2011. Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds. *Nat Rev Microbiol* 9:166–176. <https://doi.org/10.1038/nrmicro2520>.
 39. Fischetti VA. 2019. Surface proteins on Gram-positive bacteria. *Microbiol Spectr* 7(4). <https://doi.org/10.1128/microbiolspec.GPP3-0012-2018>.
 40. Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7:1511–1522. <https://doi.org/10.1038/nprot.2012.085>.
 41. Pizarro-Cerda J, Cossart P. 2006. Bacterial adhesion and entry into host cells. *Cell* 124:715–727. <https://doi.org/10.1016/j.cell.2006.02.012>.
 42. Pounton JA, Smith WD, Saalbach G, Crow A, Kehoe MA, Banfield MJ. 2010. A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. *J Biol Chem* 285:33858–33866. <https://doi.org/10.1074/jbc.M110.149385>.
 43. Ofek I, Hasty DL, Sharon N. 2003. Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol Med Microbiol* 38:181–191. [https://doi.org/10.1016/S0928-8244\(03\)00228-1](https://doi.org/10.1016/S0928-8244(03)00228-1).
 44. Asadi A, Razavi S, Talebi M, Gholami M. 2019. A review on anti-adhesion therapies of bacterial diseases. *Infection* 47:13–23. <https://doi.org/10.1007/s15010-018-1222-5>.
 45. Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 46. Bron S, Meijer WJJ, Holsappel S, Haima P. 1991. Plasmid instability and molecular cloning in *Bacillus subtilis*. *Res Microbiol* 142:875–883. [https://doi.org/10.1016/0923-2508\(91\)90068-L](https://doi.org/10.1016/0923-2508(91)90068-L).
 47. Sonhammer EL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175–182.
 48. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
 49. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* 29:2994–3005. <https://doi.org/10.1093/nar/29.14.2994>.
 50. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, Yu YK. 2005. Protein database searches using compositionally adjusted substitution matrices. *FEBS J* 272:5101–5109. <https://doi.org/10.1111/j.1742-4658.2005.04945.x>.
 51. Wright ES. 2015. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 16:322. <https://doi.org/10.1186/s12859-015-0749-z>.
 52. Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89:10915–10919. <https://doi.org/10.1073/pnas.89.22.10915>.
 53. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
 54. Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 61:1061–1067. <https://doi.org/10.1093/sysbio/sys062>.
 55. Galata V, Fehlmann T, Backes C, Keller A. 2019. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res* 47:D195–D202. <https://doi.org/10.1093/nar/gky1050>.
 56. Bullock WO, Fernandez JM, Short JM. 1987. XL1-blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with Beta-galactosidase selection. *Biotechniques* 5:376–379.
 57. Yanish-Perron C, Vieira J, Messing J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. *Gene* 33:103–119. [https://doi.org/10.1016/0378-1119\(85\)90120-9](https://doi.org/10.1016/0378-1119(85)90120-9).
 58. Singh PK, Ramachandran G, Duran-Alcalde L, Alonso C, Wu LJ, Meijer WJ. 2012. Inhibition of *Bacillus subtilis* natural competence by a native, conjugative plasmid-encoded ComK repressor protein. *Environ Microbiol* 14:2812–2825. <https://doi.org/10.1111/j.1462-2920.2012.02819.x>.
 59. Koo BM, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, Wapinski I, Galardini M, Cabal A, Peters JM, Hachmann AB, Rudner DZ, Allen KN, Typas A, Gross CA. 2017. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst* 4:291–305 e7. <https://doi.org/10.1016/j.cels.2016.12.013>.