

Structural and Biochemical Analysis of OrfG: The VirB8-like Component of the Conjugative Type IV Secretion System of ICESt3 From Streptococcus thermophilus

Julien Cappele, Abbas Mohamad Ali, Nathalie N. Leblond-Bourget, Sandrine Mathiot, Tiphaine Dhalleine, Sophie Payot, Martin Savko, Claude Didierjean, Frederique Favier, Badreddine Douzi

▶ **To cite this version:**

Julien Cappele, Abbas Mohamad Ali, Nathalie N. Leblond-Bourget, Sandrine Mathiot, Tiphaine Dhalleine, et al.. Structural and Biochemical Analysis of OrfG: The VirB8-like Component of the Conjugative Type IV Secretion System of ICESt3 From Streptococcus thermophilus. Frontiers in Molecular Biosciences, 2021, 8, 10.3389/fmolb.2021.642606 . hal-03196531

HAL Id: hal-03196531

https://hal.univ-lorraine.fr/hal-03196531

Submitted on 12 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 18 March 2021
doi: 10.3389/fmolb.2021.642606

Structural and Biochemical Analysis of OrfG: The VirB8-like Component of the Conjugative Type IV Secretion System of ICE*St3* From *Streptococcus thermophilus*

Julien Cappele[1], Abbas Mohamad Ali[2†], Nathalie Leblond-Bourget[2], Sandrine Mathiot[1], Tiphaine Dhalleine[3], Sophie Payot[2], Martin Savko[4], Claude Didierjean[1], Frédérique Favier[1] and Badreddine Douzi[2]**

[1]Université de Lorraine, CNRS, CRM2, Nancy, France, [2]Université de Lorraine, INRAE, DynAMic, Nancy, France, [3]Université de Lorraine, INRAE, IAM, Nancy, France, [4]Synchrotron SOLEIL, L'Orme des Merisiers, Saint Aubin BP 48, Gif-sur-Yvette, France

OPEN ACCESS

Edited by:
Anastassios C. Papageorgiou, University of Turku, Finland

Reviewed by:
Julien Bergeron, King's College London, United Kingdom
Zhao Wang, Baylor College of Medicine, United States

***Correspondence:**
Frédérique Favier
frederique.favier@univ-lorraine.fr
Badreddine Douzi
badreddine.douzi@inrae.fr
†Present address:
Abbas Mohamad Ali, CEACadarache/BIAM/St Paul Les Durance, France

Specialty section:
This article was submitted to Structural Biology, a section of the journal Frontiers in Molecular Biosciences

Received: *16 December 2020*
Accepted: *01 February 2021*
Published: *18 March 2021*

Citation:
Cappele J, Mohamad Ali A, Leblond-Bourget N, Mathiot S, Dhalleine T, Payot S, Savko M, Didierjean C, Favier Fé and Douzi B (2021) Structural and Biochemical Analysis of OrfG: The VirB8-like Component of the Conjugative Type IV Secretion System of ICESt3 From Streptococcus thermophilus. *Front. Mol. Biosci. 8:642606. doi: 10.3389/fmolb.2021.642606*

Conjugative transfer is a major threat to global health since it contributes to the spread of antibiotic resistance genes and virulence factors among commensal and pathogenic bacteria. To allow their transfer, mobile genetic elements including Integrative and Conjugative Elements (ICEs) use a specialized conjugative apparatus related to Type IV secretion systems (Conj-T4SS). Therefore, Conj-T4SSs are excellent targets for strategies that aim to limit the spread of antibiotic resistance. In this study, we combined structural, biochemical and biophysical approaches to study OrfG, a protein that belongs to Conj-T4SS of ICE*St3* from *Streptococcus thermophilus*. Structural analysis of OrfG by X-ray crystallography revealed that OrfG central domain is similar to VirB8-like proteins but displays a different quaternary structure in the crystal. To understand, at a structural level, the common and the diverse features between VirB8-like proteins from both Gram-negative and -positive bacteria, we used an *in silico* structural alignment method that allowed us to identify different structural classes of VirB8-like proteins. Biochemical and biophysical characterizations of purified OrfG soluble domain and its central and C-terminal subdomains indicated that they are mainly monomeric in solution but able to form an unprecedented 6-mer oligomers. Our study provides new insights into the structural analysis of VirB8-like proteins and discusses the interplay between tertiary and quaternary structures of these proteins as an essential component of the conjugative transfer.

Keywords: conjugation, Gram - positive bacteria, type IV secretion system, Integrative and Conjugative Element (ICE), VirB8-like proteins

INTRODUCTION

Conjugation constitutes one of the major mechanisms of horizontal gene transfer (Frost et al., 2005). Besides their crucial role in prokaryotic evolution and adaptation, conjugative elements including plasmids and Integrative and Conjugative Element (ICEs) constitute the most important vectors of antibiotic resistance spreading among bacteria. Conjugative transfer is ensured by dynamic multi-

subunit cell-envelope spanning nanomachines called conjugative type IV secretion systems (Conj-T4SSs) encoded by conjugative elements.

Based on the well-studied archetypal VirB/D4 Conj-T4SS system of the Ti conjugative plasmid from the Gram-negative bacterium *Agrobacterium tumefaciens*, it is commonly assumed that Vir subunits can be grouped into three classes (Cascales and Christie, 2003; Alvarez-Martinez and Christie, 2009; Fronzes et al., 2009; Grohmann et al., 2018). One encloses the energetic components VirB4, VirB11, and VirD4. VirB4 and VirB11 belong to traffic ATPases and are involved in the functioning of T4SS apparatus and in pili biogenesis whereas the coupling protein VirD4 (TC4P) is involved in DNA substrate recognition. The second class corresponds to the inner membrane platform (IMP), which is composed of the membrane proteins VirB3, VirB6, and VirB8. The IMP serves as a docking station for DNA substrate and constitutes the inner membrane channel. The third class forms the outer membrane core complex. It is composed of VirB7, VirB9, and VirB10 proteins. The outer membrane core complex forms a conducting channel that traverses the periplasmic space and the outer membrane barrier, establishing a continuity for the inner membrane channel. Other proteins are involved in the conjugative process. These proteins include the transglycosylase VirB1 that is important for the establishment of the T4SS apparatus within the peptidoglycan layer in the periplasmic space and of the extracellular pili involved in DNA transfer or in host cell attachment. Conjugative pili are predominantly composed by the VirB2 pilin and are tipped by the VirB5 adhesin (Grohmann et al., 2018; Gonzalez-Rivera et al., 2019). DNA transfer process through Conj-T4SS requires DNA recruitment and delivery to the T4SS apparatus. Initial steps of DNA recruitment are conducted by the relaxosome complex constituted by the relaxase and accessory factors that recognise and bind specifically to the origin of transfer *oriT* sequence located in the conjugative element. The nick of the *oriT* sequence is catalysed by the relaxase that binds covalently to the nascent 5' end of the DNA strand to be transferred (T-strand). Straight after, the relaxase-T-strand nucleoprotein docks to T4CP, which consequently unfolds the relaxase and ensures the transfer of the nucleoprotein to the Conj-T4SS channel. Once transferred to the recipient cell, relaxase catalyses the recircularization of the T-strand followed by the second strand synthesis (Ilangovan et al., 2017).

Conj-T4SSs in Gram-positive bacteria are more puzzling compared to those found in Gram-negative bacteria. The lack of the outer membrane components together with the incapacity to detect associated extracellular pili raise questions about the assembly mode of a conducting channel that crosses the cell-wall and ensures a cell-to-cell contact. Several lines of evidence support the idea that Conj-T4SSs from Gram-positive bacteria emerged from diderm Conj-T4SSs and form a "minimized" system spanning the cytoplasmic membrane and the murein cell wall (Bhatty et al., 2013). The best-characterized systems are those encoded by conjugative plasmids pCW3 from *Clostridium perfringens*, and pIP501 and pCF10 from *Enterococcus faecalis* (Dunny, 2007; Goessweiner-Mohr et al., 2013a; Wisniewski and Rood, 2017). Based on the structural and genetic information gathered from previous studies, it is assumed that the Conj-T4SSs from Gram-positive bacteria are essentially composed of subunits related to components recovered in Gram-negative T4SS. These ones comprise 1) VirB4 and VirD4-like energetic components, 2) VirB3-, VirB6- and VirB8-like membrane subunits that assemble in the cytoplasmic membrane to form the membrane platform, 3) VirB1-like transglycosylase and 4) surface adhesins crucial in cell-to-cell contact or cells aggregation (Bantwal et al., 2012; Arends et al., 2013; Laverde Gomez et al., 2014; Kohler et al., 2017; Dahmane et al., 2018; Kohler et al., 2018). Finally, Conj-T4SSs from Gram-positive bacteria also comprise a set of accessory proteins with undefined functions (Wisniewski et al., 2015). However, the limited structural and mechanistic information on these components restrict our understanding of their mode of assembly and function.

Despite the similitudes in the early steps of DNA processing, fundamental differences exist in terms of channel architecture between Gram-negative and -positive Conj-T4SSs. Indeed, two components, a VirB1-like transglycosylase and a VirB8-like protein were predicted to extend their C-terminal domains along the exterior face of the cytoplasmic membrane and form a protruding channel for DNA translocation through the cell wall (Bhatty et al., 2013). Unfortunately, such structural organization was not fully compatible with the crystal structures of the VirB8-like subunits TcpC and TraM from pCW3 and pIP501, respectively (Porter et al., 2012, Goessweiner-Mohr et al., 2013a). Nevertheless, is it not excluded that channel establishment depends on VirB8-like interaction with other cytoplasmic membrane subunits including the VirB1-like transglycosylase.

Two types of mobile genetic elements are autonomous for their conjugative transfer: plasmids and ICEs. With the expansion of available genomes, it emerges that ICEs are the most widespread mobile genetic elements that use conjugation to propagate autonomously intra or interspecies (Guglielmini et al., 2011; Santoro et al., 2014). The 28 kb ICE*St3* from *Streptococcus thermophilus* is the prototype of the ICE*St3*/ICE*Bs1*/Tn*916* superfamily, the most widespread superfamily among Streptococci (Ambroset et al., 2015). A large number of derivatives from ICE*St3*/ICE*Bs1*/Tn*916* contribute to the spread of various antimicrobial resistance genes in human or zoonotic pathogens including tetracycline and macrolide resistance genes (Roberts and Mullany, 2011; Schroeder and Stephens, 2016). ICE*St3* was shown to be self-transferable intra- or inter-specifically including towards the human pathogen *Streptococcus pyogenes* and the opportunistic pathogen *E. faecalis* (Bellanger et al., 2009). The ICE*St3* conjugation module encodes 14 proteins (OrfA to OrfN) (Burrus et al., 2002). Several of these proteins are homologous to the IMP components of the Gram-negative conjugative T4SS, whereas others (OrfB, E, F, H, L, M, N) are not. The presence of these latter Orfs with undefined structural homologues and functional information suggests distinct structural and assembly features of the conjugative apparatus of ICE*St3* and may reflect a specific functioning mechanism.

This study belongs to a multiapproach project that aims to understand the architecture and the assembly mode of Conj-T4SS from ICEs found in Gram-positive bacteria. Here, by using the ICE*St3* as biological model, we focus on the structure-function analysis of several proteins of the Conj-T4SS. Among them, OrfG shares a low sequence identity (22%) with TcpC, the VirB8-like component of pCW3. A structural analysis revealed that the OrfG central domain adopts a NTF2-like fold and likely presents more structural similarities with VirB8-like proteins TcpC and TraM from Gram-positive bacteria than Gram-negative VirB8 proteins. We used structural alignment tools to identify different structural classes among VirB8-like proteins. Our *in-silico* analysis revealed that Gram-positive VirB8-like proteins form a distinct class from Gram-negative ones, supporting a different evolutionary pathway. As described for soluble domains of Gram-positive VirB8-like proteins, OrfG soluble domain was found to be mainly monomeric in solution. However, a small fraction of OrfG was able to self-assemble into 6-mer oligomers. The significance of this quaternary structure is discussed in view of the role of this protein in Gram-positive Conj-T4SSs.

MATERIALS AND METHODS

Bacterial Strains, Media and Chemicals

Escherichia coli DH5α was used for cloning procedures, BL21 (DE3) was used for protein expression and production. The genotype of each strain is listed in **Supplementary Table S1**. Lysogenic Broth (LB) was used for bacterial culture. Kanamycin (50 μg/ml) was added to the culture medium.

Plasmid Construction

Plasmids used in this study are listed in **Supplementary Table S1**. Polymerase Chain Reactions (PCR) were performed using a BioRad T100 Thermocycler thermal Cycler using Phusion High-fidelity DNA Polymerase (Thermo Scientific). *Streptococcus thermophilus* LMG18311 harboring ICE*St3* genomic DNA was used as template for PCR amplifications. Custom oligonucleotides (Eurogentec) used for cloning procedures were listed in **Supplementary Table S1**. For protein expression and production, the DNA sequence encoding $OrfG_{64–331}$, $OrfG_{64–204}$, $OrfG_{64–215}$ and $OrfG_{223–331}$ were cloned into the pET28-TRX leading to pET28-TRX-$OrfG_{64–204}$, pET28-TRX-$OrfG_{64–215}$, and pET28-TRX-$OrfG_{223-331}$, respectively. For $orfG_{64–331}$ and $orfG_{223–331}$, restriction-ligation procedure was used for genes cloning. pET28-TRX-$OrfG_{64–204}$ and pET28-TRX-$OrfG_{64–215}$, plasmid derivatives were obtained by insertion of Stop codon at position 205 and 216 respectively using pET28-TRX-$OrfG_{64–331}$ as template.

Production and Purification of $OrfG_{64–331}$, $OrfG_{64–204}$, $OrfG_{64–215}$, and $OrfG_{223–331}$

E. coli BL21 (DE3) cells carrying the pET28-TRX derivatives were grown in LB supplemented with Kanamycin (50 μg/ml), at 37°C to an OD_{600} ~ 0.5. Expression of the constructions was then induced by addition of IPTG (0.5 mM) and cultures were pursued for 18 h at 25°C. Cells were harvested, resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM EDTA, 50 μg/ml lysozyme, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and submitted to five cycles of sonication. After addition of 20 μg/ml DNase and 20 mM $MgCl_2$, the soluble fraction was obtained by centrifugation for 40 min at 16,000 × *g*. Recombinant proteins were purified by ion metal affinity chromatography using a 5-ml Nickel HisTrap™ HP Column on an ÄKTA prime apparatus (GE healthcare) pre-equilibrated in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM Imidazole (buffer A). After several washes in buffer A, 6×His tagged proteins were eluted in buffer A supplemented with 250 mM Imidazole, cleaved with the 1 mg/ml TEV protease and dialyzed against 50 mM Tris-HCl, 300 mM NaCl pH 8.0 for 18 h at 4°C and loaded onto a HisTrap™ HP pre-equilibrated in buffer A, which selectively retains the TEV protease, the uncleaved 6×His-tagged fusion proteins and contaminants. The native proteins were collected in the flow-through, concentrated on Centricon (Millipore; 10 kDa cutoff), and passed through a Sephadex 200 26/60 column pre-equilibrated with 50 mM Tris-HCl pH 8.0, 100 mM NaCl or 20 mM HEPES pH 8.0, 100 mM NaCl.

Crystallization and Structure Determination

Crystallization was conducted using the sitting drop, vapor diffusion method at 293 K. $OrfG_{64–204}$ crystals were obtained by mixing 0.3 μl of protein (35 mg/ml, 100 mM Tris-HCl pH 8.0) with 0.3 μl of the reservoir solution containing 2 M ammonium sulfate, 100 mM Tris-HCl, pH 8.5. All drops were prepared by the crystallization robot Oryx8 (Douglas Instruments, Hungerford, United Kingdom) in 96 well 2-lens MRC plates (SWISSCI AG, Neuheim, Switzerland). Crystals reached their maximal size after approximatively 1 week. To ensure phasing, osmium (IV) hexachloride di-potassium salt was added to crystal-containing drops (10 mM final concentration) and left for 24 h. Crystals were then mounted on nylon loops and quickly back-soaked into their reservoir solution +20% glycerol and immediately flash frozen in liquid nitrogen to be stored until data collection.

All X-ray diffraction experiments were carried out at Synchrotron SOLEIL (St-Aubin, France) on Proxima-2A beamline (**Table 1**). Measurements were performed at 100 K. A first crystal without osmium was used to collect a native dataset at $\lambda = 0.9801$ Å, up to a resolution of 1.59 Å. At an advanced stage of the refinement the Fourier difference map remained noisy. The associated R_{free} factor was 21.2% (42.8% in the highest resolution shell). Thus, we decided to limit the resolution to 1.75 Å (data processed with XDS (Kabsch, 2010) and scaled with Aimless (Evans and Murshudov, 2013)). From an osmium containing crystal, eight datasets (400°span, 4,000 images each) were collected at osmium peak absorption $\lambda = 1.1397$ Å, at different kappa angles. Merging all of them was mandatory to get a final data set at 2.0 Å with sufficient quality for successful phasing (data processed with XDS and XSCALE). Preliminary phases were determined using the SAD Pipeline implemented in PHASER CCP4i (McCoy et al., 2007) and the partial results were enhanced by ARP/wARP (Langer et al., 2008). The resulting model was further improved with the native dataset at 1.75 Å, using PHENIX (Liebschner et al., 2019) and COOT (Emsley et al.,

TABLE 1 | Statistics of data collection for the native crystal and for the osmium derivative crystal, and refinement of $OrfG_{64–204}$ model.

	Native	Osmium derivative
Data collection		
Space group	$P4_12_12$	$P4_12_12$
Beamline	Proxima-2A, SOLEIL	Proxima-2A, SOLEIL
Unit cell parameters (Å, °)	50.2 50.2 120.9 90. 90. 90.	50.4 50.4 121.3 90. 90. 90.
Wavelength (Å)	0.9801	1.1398
Resolution (Å)[a]	46.33–1.75 (1.78—1.75)	46.62–2.00 (2.05—2.00)
Observations[a]	411,273 (20,684)	1,637,839 (40,984)
Unique reflections	16,372 (1,116)	20,272 (1,494)[b]
R_{merge}[a]	0.027 (0.499)	0.107 (1.028)
R_{meas}[a]	0.028 (0.514)	0.107 (1.048)
$CC_{1/2}$	1.0 (0.975)	1.0 (0.845)
Mean I/σ[a]	67.5 (7.3)	55.6 (2.8)
Completeness (%)[a]	100. (100.)	100. (99.4)
Anomalous completeness (%)[a]	—	100. (98.6)
Multiplicity[a]	25.1 (23.8)	80.1 (27.4)
Mean anomalous difference ΔF/σ[a]	—	1.21 (0.74)
Figure of merit (/)[c]	—	0.13/0.37
First model[d]	—	4 chains, 63 aa, 31 sequenced
Model refinement		
Resolution (Å)[a]	46.33–1.75 (1.80–1.75)	
No. of protein residues	101 (3 chains)	
No. of non-hydrogen protein atoms	842	
No. of waters	81	
R_{cryst}[a]	0.200 (0.294)	
R_{free}[ae]	0.208 (0.427)	
RMSD bonds (Å)	0.014	
RMSD angles (°)	1.635	
Ramachadran plot[f]		
Favored (%)	100	
Outliers (%)	0.0	
Rotamer outliers (%)	1.1	
Clashscore[f]	5.42 (94)	
MolProbity score (PR)[f]	1.32 (98)	
Average B (Å^2)	35.4	
Protein	33.0	
Water	42.1	
Coord. error (Å)	0.10	
PDB code	6zgn[g]	

[a]*Values for the outer shell of data are given in parentheses.*
[b]*Number of reflections in the data set where the Friedel mates are not merged.*
[c]*Before/after solvent flattening; the same values were obtained for both hands.*
[d]*As built by Buccaneer (CCP4i); no model was obtained for the other hand.*
[e]*Five percent test set.*
[f]*From Molprobity. The percentile ranks (PR) for Clashscore and MolProbity score are given in parentheses.*
[g]*PDB validation report is available in supplementary materials.*

2010) (**Table 1**). The asymmetric unit contained a single $OrfG_{64–204}$ monomer. Several regions were unobserved in the electron density when contoured at 1.2 σ, especially at the N-terminal end. These regions were assumed to be disordered and were not included in the model. The final model contained residues Asp92-Val133 + Val139-Ile168 + Glu176-Ala204, i.e., 70% of the 143 amino acids of the $OrfG_{64–204}$ subdomain. It is worth noting an important blob of density in the pocket delineated by Tyr106-Phe107, Arg140-Ser143, Tyr164, Glu166 and Phe202 (**Supplementary Figure S1**). This blob could be assigned to none of the compounds used for protein crystallization. It probably resulted from a significant affinity for a ligand present in the protein expression cells or in the successive media used to prepare the protein sample. The final $OrfG_{64–204}$ structure was deposited in the Protein Data Bank with ID 6zgn. All three-dimensional structure representations were obtained using PYMOL (Schrödinger, LLC).

In Silico Analysis, Sequence Alignment, Structure Comparison and Interface Analysis

Protein sequences were retrieved from NCBI and compared using clustalW from MEGA7 (Kumar et al., 2016). Phylogenetic trees were built with MEGA7 by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993) without outgroup.

TMD prediction was conducted using CCTOP (http://cctop.enzim.ttk.mta.hu/). PSIPRED (Buchan and Jones, 2019) and HHPRED (Zimmermann et al., 2018) were used for OrfG secondary structures and subdomains structures predictions, respectively. COILS (Lupas et al., 1991), MULTICOIL (Wolf et al.,

1997) and DEEPCOIL (Gabler et al., 2020) were used to examine the presence of a coiled-coil domain on OrfG.

The first step of structure comparison was conducted with both DALI (Holm and Sander, 1995) and PDBefold (Krissinel and Henrick, 2004) to fetch a complete list of VirB8 structures (26 PDB entries). From them, a non-redundant set of 16 structures was conserved (2cc3, 3ub1, 3wz3, 3wz4, 4akz, 4ec6, 4jf8, 4kz1, 4lso, 4mei, 4nhf, 4o3v, 5aiw, 5cnl, 5i97, 6iqt) and added to the new $OrfG_{64–204}$. Then the mTM-align server (Dong et al., 2018a) based on the robust TM-Align method (Dong et al., 2018b) performed the structural alignment and produced the structure-based "phylogenetic tree." PyMOL and an associated homemade python script were used to emphasize visually structure differences and similarities (available upon request). Briefly, a color gradient was applied to each position depending on the rmsd value calculated on Cα carbons of structurally aligned amino acids at this position. Conserved or similar residues were also automatically highlighted. Furthermore, the superposition of all VirB8 monomers produced by the mTM-align server was used to compare oligomers interfaces and to visually identify interesting features.

Biophysical Characterization of Native and Truncated $OrfG_{64–331}$ Proteins

To determine the molecular weight of the purified proteins, size exclusion chromatography (SEC) coupled to MALS analysis was performed on MiniDAWN TREOS II coupled to Superdex 200 10/300 increase column (GE Healthcare) mounted on FPLC system (AKTA purifier). We used 20 mM Tris-HCl pH 8.0, 100 mM NaCl buffer at 0.5 ml/min flow.

In Vitro Chemical Crosslinking

Crosslinking experiments were performed for $OrfG_{64–331}$, $OrfG_{64–204}$ and $OrfG_{223–331}$ obtained after SEC purification and present in 20 mM HEPES pH 8.0, 100 mM NaCl buffer. The used protocol is described as follow: 20 µM of each protein were mixed or not with increasing concentration of paraformaldehyde (PFA) from 0.05 to 5% in incubated for 1 h at 37°C. The crosslinking reaction was stopped by the adding of 1 M Tris-HCl at a final concentration of 80 mM and the mixture was incubated for 10 min at RT. To examine crosslinked products, 15 µl of SDS-PAGE loading buffer (4×) were added and 10 µl from each reaction was analyzed by 12% SDS-PAGE for $OrfG_{64–331}$ and 15% SDS-PAGE for $OrfG_{64–204}$ and $OrfG_{223–331}$.

RESULTS

OrfG Belongs to the TcpC Family

Previous studies showed that the integrative and conjugative element ICE*St3* from *S. thermophilus* is an autonomous element able to self-transfer by conjugation intra and interspecifically (Bellanger et al., 2009). Bioinformatic analyses of ICE*St3* predicted the presence of a conjugation module that harbors 14 genes. Among them, *orfJ* was recently showed to encode the relaxase that belongs to the MobT superfamily (Soler et al., 2019). In this study, we focused on the structural and biochemical studies of a putative transfer protein OrfG. Blastp analysis of OrfG indicated that this protein is composed of a TcpC domain (pfam12642) from position 98–316 (e-value $7.4e^{-39}$) (**Figure 1A**). This domain is homologous to that of several TcpC conjugal transfer proteins encoded by diverse ICEs or conjugative plasmids, including the VirB8-like subunit TcpC from pCW3. The archetypal TcpC from pCW3 is a membrane protein that was shown to be essential for the conjugation of pCW3 from *C. perfringens* (Porter et al., 2012). TcpC consists of an N-terminal cytoplasmic domain followed by a transmembrane domain (TMD) which encompasses residues 57–79 and a soluble domain extended in the cell-wall. Despite a poor sequence identity/similarity between OrfG and TcpC (23% identity/38% similarity) in their TcpC domain, *in silico* analysis using PSIPRED, HHPRED and CCTOP revealed that OrfG shares the same domain organization as observed for TcpC. Thus, OrfG is composed of an N-terminal tail followed by a TMD (residues 37–59) and a soluble domain that we called $OrfG_{64–331}$ (**Figure 1A**)

Our phylogenetic analysis of OrfG and TcpC orthologs indicated that OrfG is more closely related to the VirB8-like protein (conB) from ICE*Bs1* (Leonetti et al., 2015) than to the other TcpC proteins (including TcpC from pCW3) (Porter et al., 2012). This is consistent with the existence of at least two sub-families of TcpC proteins (**Figure 1B**). Given the similarities in domain organization between OrfG and TcpC, their distant evolutionary relationship motivated us to focus on the structural and biochemical characterization of OrfG.

OrfG Central Domain ($OrfG_{64–204}$) Displays a VirB8-like Structure

Owing to the lack of structural data on Gram-positive Conj-T4SS components especially those involved in ICE conjugation, we envisaged structure determination of OrfG by X-ray diffraction. For this purpose, we purified to homogeneity a large amount of the OrfG soluble domain ($OrfG_{64–331}$: from Ser64 to Asp331). Despite an easy growing of crystals in multiple and different conditions, their poor X-ray diffraction power prevented the acquisition of data of sufficient quality. To overcome these technical issues, and based on the secondary structure prediction of OrfG as well as the sequence alignment between OrfG and TcpC, we chose to express and purify to homogeneity the separate central ($OrfG_{64–204}$: from Ser64 to Ala204) and C-terminal ($OrfG_{223–331}$: from Ala223 to Asp331) subdomains (**Figure 1A**). Crystallization trials conducted on both forms resulted in suitable tetragonal crystals for $OrfG_{64–204}$ while no positive result was obtained for $OrfG_{223–331}$. Experimental phases were determined by using a single-wavelength anomalous dispersion method (Porter et al., 2012) on crystals soaked with an osmium derivative. A native crystal allowed final model refinement to 1.75 Å resolution (R_{free} = 20.8%). Detailed data and refinement statistics are given in **Table 1**. Note that the electron density of OrfG central domain displayed a blob (**Supplementary Figure S1**) that remained unassigned despite attempts to fit various compounds in it. This density was located between β_4 and the C-terminal end of α_1.

The polypeptide chain adopts an alpha + beta fold, composed of two α-helices named α_1 and α_2, followed by a β-sheet of four

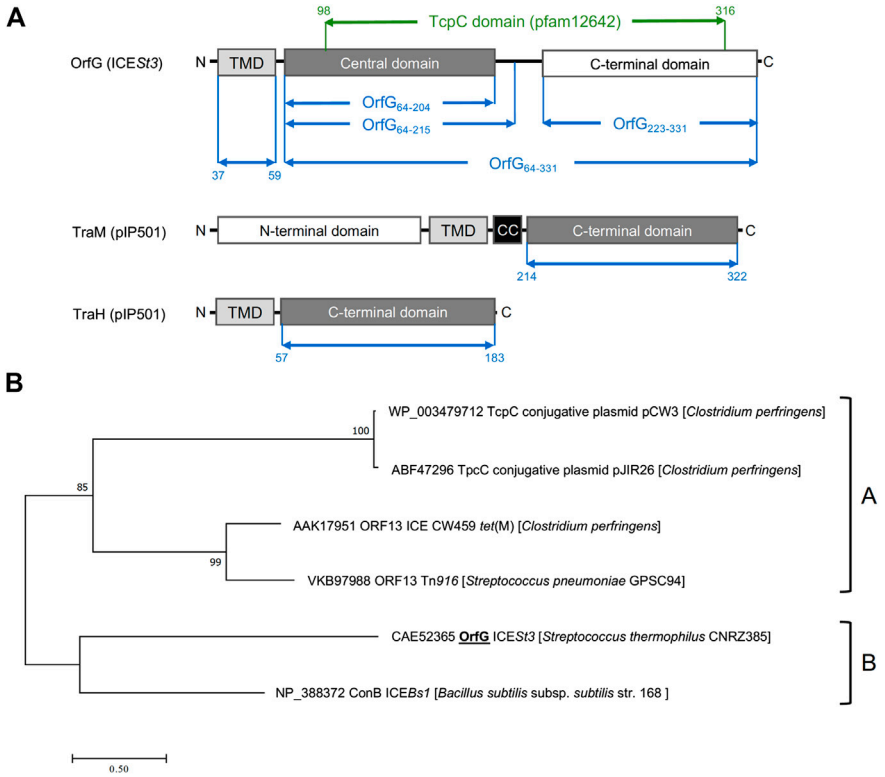


FIGURE 1 | OrfG belongs to the TcpC superfamily **(A)** Schematic representation of OrfG, TraM and TraH subdomains organization. Boundaries of OrfG subdomains and TraM and TraH NTF2-like domains are stated. TMD for transmembrane domain; CC for coiled-coil domain **(B)** Phylogenetic tree of representative members from the TcpC superfamily including OrfG. This tree was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Guindon et al., 2010). The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured as the number of substitutions per site. All positions containing gaps and missing data were eliminated. There was a total of 284 positions in the final dataset. Evolutionary analyses were conducted using MEGA7 (Tamura and Nei, 1993).

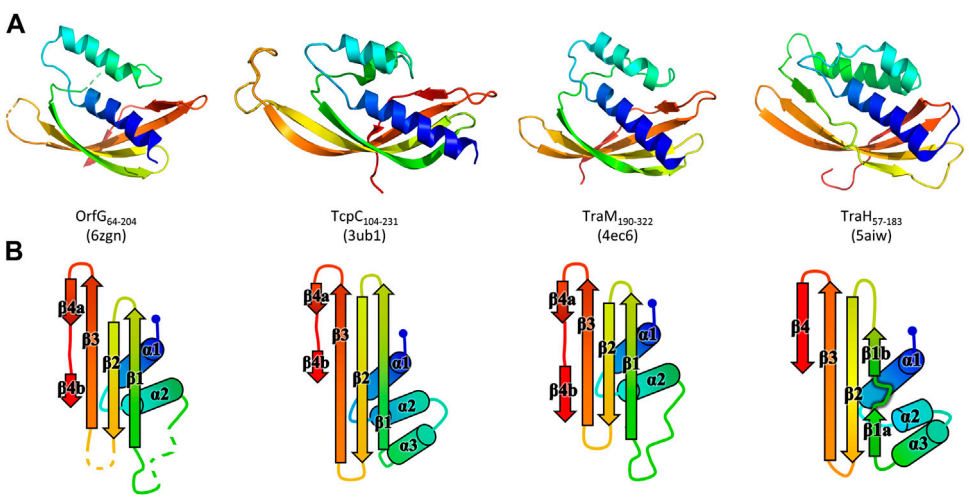


FIGURE 2 | Structures of $OrfG_{64–204}$, $TcpC_{104–231}$, $TraM_{190–322}$, and $TraH_{57–183}$. From left to right respectively, the four monomers (PDB id 6zgn, 4ec6, 3ub1, and 5aiw) are displayed in the same orientation and represented as rainbow-colored cartoons, with the N-terminal end of the polypeptide chain in blue and the C-terminal end in red **(A)**. Their topology is schematized with the same rainbow-color code and the name of the secondary structures **(B)**.

strands (β_1–β_4), the last of which has a marked disruption into two subparts β_{4a} and β_{4b} (**Figure 2**). Cross-comparisons performed by DALI (Holm and Sander, 1995) and PDBefold (Krissinel and Henrick, 2004) classify the spark OrfG central domain in the NTF2-like superfamily. Its folding pattern mainly consists in a highly curved antiparallel beta sheet that wraps around a central alpha helix surrounded by shorter additional ones. As expected from sequence analysis, the closest VirB8-like structural neighbours of $OrfG_{64–204}$ is the TcpC central domain ($TcpC_{104–231}$, i.e., the first domain of $TcpC_{99–359}$, PDB entry 3ub1 (Porter et al., 2012)), with an rmsd of 1.77 Å for 92 aligned amino acids. This search also highlighted the structure of the C-terminal domain of TraM from pIP501 ($TraM_{190–322}$, PDB entry 4ec6 (Goessweiner-Mohr et al., 2013b), rmsd of 1.90 Å for 96 aligned amino acids) as a close structural homolog (**Figure 2**). The folds of $TcpC_{104–231}$, $TraM_{190–322}$ and $OrfG_{64–204}$ differ by the length of their secondary structures and connecting loops. Furthermore, $TcpC_{104–231}$ has an additional helix α_3 prior to the β-sheet. Both TcpC and TraM were described as Gram-positive equivalents of VirB8 from VirB/D4 T4SS from *A. tumefaciens*. Despite their low sequence identity and their different domain organization, it was proposed that proteins from Gram-positive Conj-T4SSs that adopt a NTF2-like fold form the "VirB8-like" family (Ambroset et al., 2015). This family also includes the soluble domain of the conjugative component TraH from pIP501 ($TraH_{57–183}$, PDB entry 5aiw (Fercher et al., 2016)) (**Figure 2**). Since the structure of $OrfG_{64–204}$ belongs to the NTF2-like superfamily as well, we propose that OrfG from ICE*St3* is a new member of the VirB8-like family found in Gram-positive Conj-T4SS.

Structural Comparison of Available VirB8-like Structures

Unexpectedly, known structures of VirB8-like proteins show subdomains that harbor an overall fold similar to the unique NTF2-like domain of Gram-negative VirB8s. This observation sparked our interest in a structural comparison of these proteins. A previous study using sequence analysis, secondary structure prediction and domain composition of a large set of sequences distinguished three classes of proteins (Goessweiner-Mohr et al., 2013b). Here we conducted a distinct analysis that focused on the available three-dimensional structures of the NTF2-like domain of these proteins, with the aim to point out the shared features and the specific characteristics of both Gram-positive and Gram-negative proteins. It was based on the superposition of their atomic coordinates, independently of their amino acid sequence. The results emphasized a surprising diversity despite the simplicity and small size of the NTF2-like fold.

The set of Gram-negative VirB8 structures encompasses 13 independent entries in the protein database (Protein Data Bank). That of known Gram-positive VirB8-like structures is poorer: the crystallographic analysis of $OrfG_{64–204}$ brings their number to four. The NTF2-like pattern is globally conserved in the 17 structures. On average, their fold consists in three antiparallel alpha helices followed by an antiparallel four-stranded sheet, for which we propose the nomenclature $\alpha_1\alpha_2\alpha_3\beta_1\beta_2\beta_3\beta_4$. However, the length of the secondary structures (especially beta strands) and their relative orientation, the curvature of the beta sheet, plus possible decorations inserted in the minimal fold bring enough diversity to make the structural superimposition delicate. We found the mTM-align algorithm as the most suited one to deal with these differences. The global superimposition of the 17 structures and the corresponding structural alignment provided by mTM-align were analyzed using an in-house developed script to allow their visualization in Pymol (*see Materials and Methods*) (**Figure 3**). The structure-based tree obtained using mTM-align indicated that Gram-positive VirB8-like structures clearly separate from the group formed by the Gram-negative ones (**Figure 4**). The presence of clusters with members that share high sequence similarities and others where modest to low sequence similarities are observed confirms that the structural alignment was performed independently from primary sequences (**Supplementary Table S2**). At least three classes appear in the Gram-negative VirB8s (named I^-, II^- and III^- for the convenience of their description) and one in the Gram-positive VirB8-like structures (I^+), while the three other proteins did not belong to identified classes (**Figure 4**). In the Gram-negative group, the three classes differ by the way their helix α_3 is disrupted in two parts α_{3a} and α_{3b}, separated by a protruding loop of four residues with two distinct conformations in classes I^- and II^-, or by a strong kink in class III^- (**Figure 4**). This disruption of α_3 is not a distinctive trait of the Gram-negative proteins since CagV from *Helicobacter pylori* (PDB entry 6iqt (Wu et al., 2019)), not assigned to a class, has a straight helix. All members of classes I^- and II^- have a small additional helix α_4 between β_3 and β_4, which is involved in protein dimerization, while members of class III^- have a loop instead. Furthermore, class III^- and CagV display a bulge that separates β_1 in two parts β_{1a} and β_{1b} (**Figure 4**).

Strikingly, the four Gram-positive VirB8-like proteins have no specific structural feature that unambiguously distinguishes them from the Gram-negative ones. However, the overall structure superposition separated them in the structural tree, which seems to be explained by the small differences disseminated along their polypeptide chains (**Figure 4**). Indeed, $OrfG_{64–204}$ has no α_3 (replaced by an unobservable region) and a strand β_4 that clearly splits in two short portions β_{4a} and β_{4b} separated by a bulge that ends with Pro200 (**Figure 2**). By comparison, $TraM_{190–322}$ (PDB entry 4ec6) has a shorter α_2, also misses α_3 and has a slightly longer loop between β_1 and β_2. In contrast, $TcpC_{104–231}$ (PDB entry 3ub1) possesses α_3 and has much longer strands β_1, β_2, and β_3, together with a longer loop between them. $TraH_{57–183}$ is the most distant structure, as it is the only one with a continuous strand β_4 with no bulge insertion, while on the contrary β_1 possesses two parts β_{1a} and β_{1b} with a bulge in between (PDB entry 5aiw) (**Figure 2**), more similar to some Gram-negative proteins (class III^- and CagV). Furthermore, TraH has a pronounced helix α_3 with four turns and a long loop towards β_1.

The most distinguishable difference between Gram-negative and Gram-positive VirB8 proteins lies in the length of the connection between β_3 and β_4. It has 10–17 amino acids in Gram-negative VirB8s (including α_4 in classes I^- and II^-), and

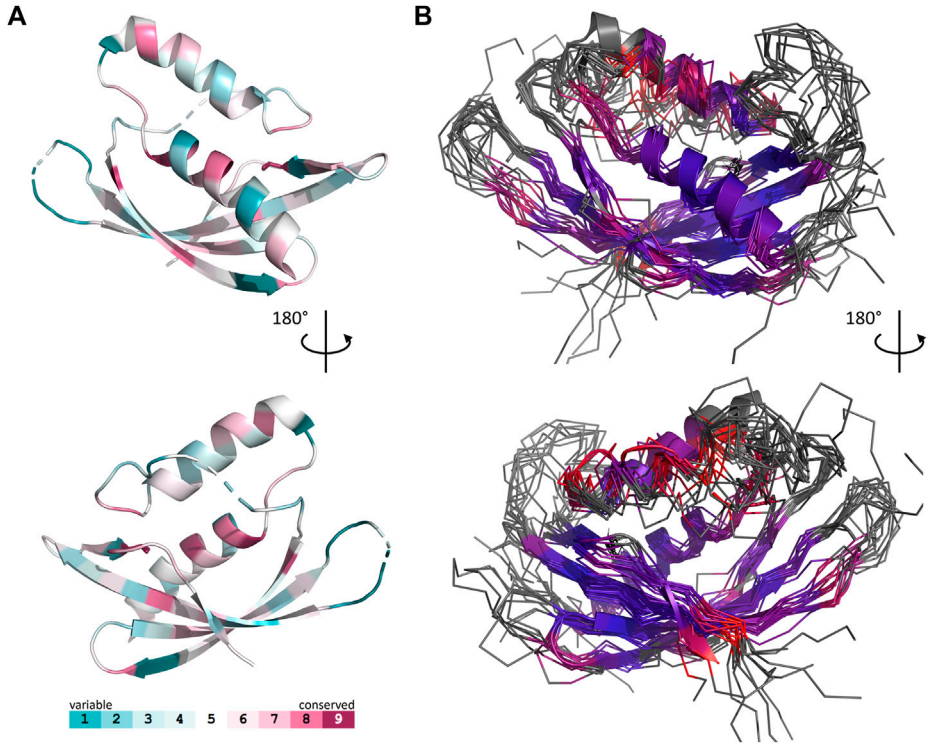


FIGURE 3 | Comparison of 17 representative three-dimensional structures of VirB8 and VirB8-like proteins. The structure of $OrfG_{64–204}$ (PDB id 6zgn) is represented as a cartoon **(A)**, with its amino acids colored according to the Consurf server (Ashkenazy et al., 2016) which analyzed the similarity calculated in the sequence alignment deduced from the structure superposition produced by mTM-Align (Dong et al., 2018a) for a set of sventeen selected proteins. The amino acid shown with sticks (Val191 in OrfG) and localised in β_4 facing α_1 is the only one whose side chain shares a similar nature in all structures, while there is no conserved residue. The 17 structures (PDB ids 2cc3, 3ub1 (limited to $TcpC_{104–231}$), 3wz3, 3wz4, 4akz, 4ec6, 4jf8, 4kz1, 4lso, 4mei, 4nhf, 4o3v, 5aiw, 5cnl, 5i97, 6iqt, 6zgn) are represented **(B)** as superposed wires (except OrfG as a cartoon) using a color-code based on the quality of the superposition of these structures. For each position, a gradient from violet to red is used depending on the low to high values of the root mean square calculated on the distance between all the possible pairs of the 17 superposed Cα carbons at this position (grey means that mTM-Align found at least one structure impossible to align). This representation highlights the resemblances and differences in the NTF2-like folds of these proteins.

only four to seven in Gram-positive ones. However, deleting all these atoms in the seventeen structure files prior to calculation of the structural tree resulted in the same classification, with only minor modifications. Thus, it seems that Gram-positive and -negative VirB8 and VirB8-like proteins separate according to details evenly distributed across the structures instead of localized marked traits that would characterize each group.

The script that we developed to analyse the structural alignment of the seventeen structures easily identified the conserved or similar residues within the different sets of proteins that we considered. Interestingly, we found no conserved residue common to all 17 structures. In the subset of the 13 Gram-negative VirB8 structures, once again, no amino acid is conserved. On the contrary, up to 30 residues are conserved within each Gram-negative class, among which many form hydrophobic contacts where the β-sheet faces α_1 (**Supplementary Figure S2**). By comparison, in the four Gram-positive VirB8-like structures, only six residues share similar natures and constitute a hydrophobic patch within the core of the protein near the C-terminal side of α_1. In addition, it is worthy to notice the presence of a conserved glutamate residue (Glu154 in OrfG) at the end of β_1, with its side chain towards the surface. Its conservation might be the result of happenstance due to the low number of analyzed structures and no apparent involvement in any observed assemblies. As for Gram-negative VirB8s, only a subset of Gram-positive VirB8-like proteins display more conserved residues (designated as Class I^+) (**Supplementary Figure S2**). The absence of any signature was already noticed in the NTF2-like superfamily (Eberhardt et al., 2013). However, it is surprising that, despite probably having the same function, this observation also applies to the restricted subset of NTF2-like superfamily proteins that constitute the VirB8-like proteins.

Crystal Packing of VirB8-like Proteins Raises the Question of their Multimerization

In the tetragonal crystal, the $OrfG_{64–204}$ monomer interacts with symmetry-related partners, among which one forms a questionable pair with a nice arrangement of the facing alpha helices (**Figure 5A**). However, the interaction involves a low surface area of 1,600 $Å^2$, or only 12% of the solvent accessible surface, as determined by PISA (Krissinel and Henrick, 2007), which rejects the arrangement as an effective quaternary structure. As well, we struggled to find an explanation to the resulting location of both N-termini in this

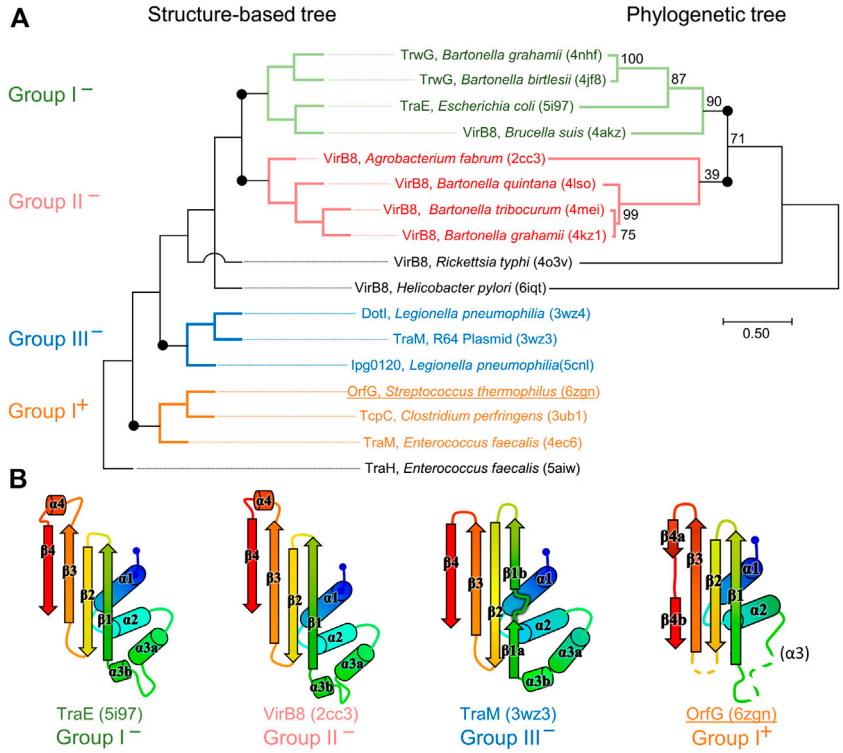


FIGURE 4 | Structural and phylogenetic trees of Gram-negative VirB8 and Gram-positive VirB8-like proteins. **(A)** Structure-based tree derived from the optimized superposition of the atomic coordinates **(left)**. 17 representative structures were retrieved from the PDB and compared by mTM-align (Dong et al., 2018a). The name of the protein, its source and PDB id are indicated for each protein. Members of class I$^-$ are in green, class II$^-$ in red, class III$^-$ in blue, and class I$^+$ in orange where the ($^-$) and ($^+$) correspond to VirB8-like structures from Gram-negative and -positive bacteria, respectively. **(B)** Secondary structure topology observed within each class is schematized at the bottom of the figure. Among the proteins compared in the structural analysis, all the sequences that belong to the PFAM group PF04335 were subjected to multiple sequence alignment and neighbor-joining tree building. The result is shown on the right-hand side of the figure. Numbers at nodes indicate the bootstrap values as percentage (1,000 replicates). Scale bar indicates the number of amino acid differences per site.

association. Indeed, it would place the transmembrane domain of each OrfG monomer at opposite sides of the dimer, but this argument should be used with caution since 27 N-terminal residues are missing in the model with respect to the crystallized protein. Altogether, the pair of $OrfG_{64-204}$ monomers observed in the crystal is not a convincing proof of the existence of a dimer with a biological relevance.

Analysis of structural similarity positioned $OrfG_{64-204}$ in the same class as $TraM_{190-322}$ and $TcpC_{104-231}$, while $TraH_{57-183}$ is more distant (**Figure 4**). The proximity of these three proteins raises once again the question of a possible multimerization of OrfG. Indeed, both $TraM_{190-322}$ and $TcpC_{99-359}$ form trimers in their crystal assembly (**Figures 5B,C**). Yet, $TraM_{190-322}$ corresponds to a single domain, like the fragment of OrfG that has been crystallized ($OrfG_{64-204}$), while $TcpC_{99-359}$ has two subdomains equivalent to the central and C-terminal domains of OrfG ($OrfG_{64-331}$, no structure available). Do the protein interfaces share homology that allows prediction of an equivalent assembly of OrfG? Alternatively, does the latter has specificities that would result in a different oligomeric form? Comparison of the trimers formed by $TraM_{190-322}$ and $TcpC_{99-359}$ emphasizes a surprising similarity in the positioning of their common domain (**Figures 5B,C**). $TraM_{190-322}$ and the central domain of TcpC ($TcpC_{104-231}$) bring their N-termini near the 3-fold axis that defines the trimer. Interactions that build this assembly are mainly located at the α_1 N-terminal moiety of one partner and near the β_1 C-terminal end of the facing partner, in a circular pattern of interactions. Detailed analyses are provided in the papers that describe these crystallographic structures (Porter et al., 2012, Goessweiner-Mohr et al., 2013b). Briefly, $TraM_{190-322}$ hides 14% of the calculated solvent-accessible surface area of one isolated monomer in this contact, which involves three hydrogen bonds and several van der Waals contacts (**Figure 5B**). Consistently with this observation, the weakness of this interaction gives little credibility to the existence of a trimer in solution. However, its existence in the crystal makes sense with respect to the prediction of a coiled-coil motif formed by some thirty residues that precede the NTF2-like domain, which could stabilize a trimeric assembly of TraM (Goessweiner-Mohr et al., 2013b). In the case of $TcpC_{99-359}$, the equivalent interface between the central domains $TcpC_{104-231}$ is a bit wider and based on more hydrogen bonds (a total of eight) (Porter et al., 2012) (**Figure 5C**). The striking observation that arises from a precise comparison of these interfaces is their total lack of commonality, except for the region they concern. Two

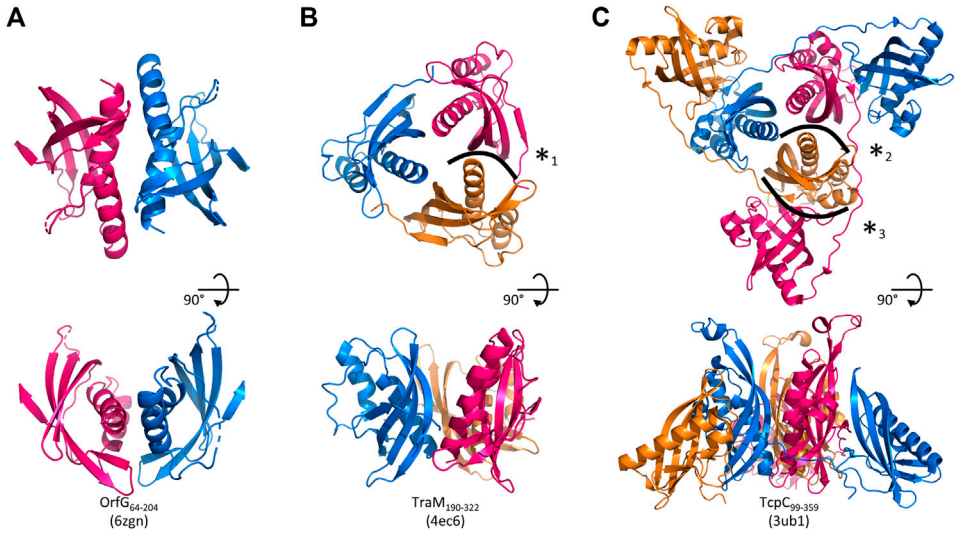


FIGURE 5 | Assemblies of Gram-positive VirB8-like proteins observed in the crystals. A crystallographic 2-fold axis relates two monomers of $OrfG_{64–204}$ (PDB id 6zgn) **(A)**, while $TraM_{190–322}$ (4ec6) **(B)** and $TcpC_{99–359}$ (3ub1) **(C)** form trimers. Each assembly is shown perpendicular **(top)** and parallel **(bottom)** to its symmetry axis. The interface with the highest surface area, here represented for $OrfG_{64–204}$, is resolutely rejected by PISA (Krissinel and Henrick, 2007). The interface (*1) in $TraM_{190–322}$ extends on ~300 $Å^2$ and involves two to nine hydrogen bonds or salt bridges depending on the observed monomer (as calculated by PISA for residues 214 to 318 of 4ec6). The equivalent interface in $TcpC_{99–359}$ (*2) represents ~430 $Å^2$ (PISA on residues 104–228 of 3ub1) and six to nine hydrogen bonds or salt bridge depending on the monomer. It is by far slighter than the interface (*3) observed between the central domain ($TcpC_{104–231}$) of one monomer and the C-terminal domain ($TcpC_{239–359}$) of a second one, which extends on ~980 $Å^2$ and involves 13–14 hydrogen bonds or salt bridges depending on the observed monomer. The structure-based sequence alignment of OrfG, TcpC and TraM available in **Figure 6** shows residues involved in these interactions.

particular examples illustrate this conclusion. First, in $TraM_{190–322}$, the side chain of Lys217 at the beginning of α_1 interacts with Tyr269 in β_1 of the partner, via van der Waals contact with its aromatic ring and the addition of a hydrogen bond with its main chain on four of the six instances observed in the asymmetric unit. At the equivalent positions of $TcpC_{99–359}$ (**Figure 6**), Gln108 and especially Phe161 have different orientations so that instead Gln108 interacts with Glu106 at the beginning of α_1 in the neighbouring monomer. The second example concerns Tyr224 (α_1) in $TraM_{190–322}$, the side chain of which forms a hydrogen bond with Asn285 (β_2) of the associated monomer. In $TcpC_{99–359}$, the residue Glu115 at the equivalent position in α_1 prefers an interaction with Ser159 in β_1 of the facing monomer. As a conclusion, the driving force that governs the trimer formation does not originate from conserved residues at this interface. However, the observation of similar multimeric forms with similar arrangements is suggestive of their probable relevance. Structure and sequence comparisons with $OrfG_{64–204}$ revealed no conservation of interacting pairs of residues (neither with $TcpC_{99–359}$ nor with $TraM_{190–322}$) but they also revealed no element that would prevent formation of the same interaction (**Figure 6**).

No trimer is observed in the crystal form of $OrfG_{64–204}$. Deletion of the C-terminal domain is a plausible explanation but it is not appropriate for $TraM_{190–322}$. A closer analysis of the different structures shows that $TraM_{190–322}$ has a C-terminal extension of five residues after β_{4b}, which follows the same direction as the first residues of the linker between the two domains of $TcpC_{99–359}$. Its penultimate residue Phe321 positions its side chain in a hydrophobic environment formed by several aromatic residues in α_1, α_2 and β_{4a} of the facing monomer. In $TcpC_{99–359}$, the linker residue Lys233 also contributes to stabilization of the trimer, but via a hydrogen bond with the polypeptide main chain following α_2. In OrfG, the equivalent Gln209 could play the same role (**Figure 6**). In order to evaluate the contribution of this extension, we solved the structure of $OrfG_{64–215}$. However, no change was induced, neither in the crystallization conditions nor in monomer packing in the crystal (data not shown) demonstrating that the linker was not sufficient to promote $OrfG_{64–215}$ multimerization.

The crystal structure of TcpC soluble domain ($TcpC_{99–359}$) shows that its C-terminal domain ($TcpC_{239–359}$) also belongs to the NTF2-like superfamily (Porter et al., 2012). It has a peripheral position in the trimeric assembly, far from the N-terminal domain ($TcpC_{104–231}$) to which it is connected through an extended coil that allows to the three monomers to intertwine (**Figure 5C**). This organization mode results in a large interface (1000 $Å^2$) between the central domain of one monomer and the C-terminal domain of the partner. Thirteen hydrogen bonds are formed, together with hydrophobic interactions. It alone accounts for much more than the interactions formed between the central domains and described above. However, this convincing arrangement, which buries 24% of the accessible surface area of each monomer, was not observed in solution (Porter et al., 2012). The authors proposed that the deleted N-terminal residues, in particular a transmembrane domain, could explain this discrepancy. The similarity between OrfG and TcpC in their domain organization and in the structure of

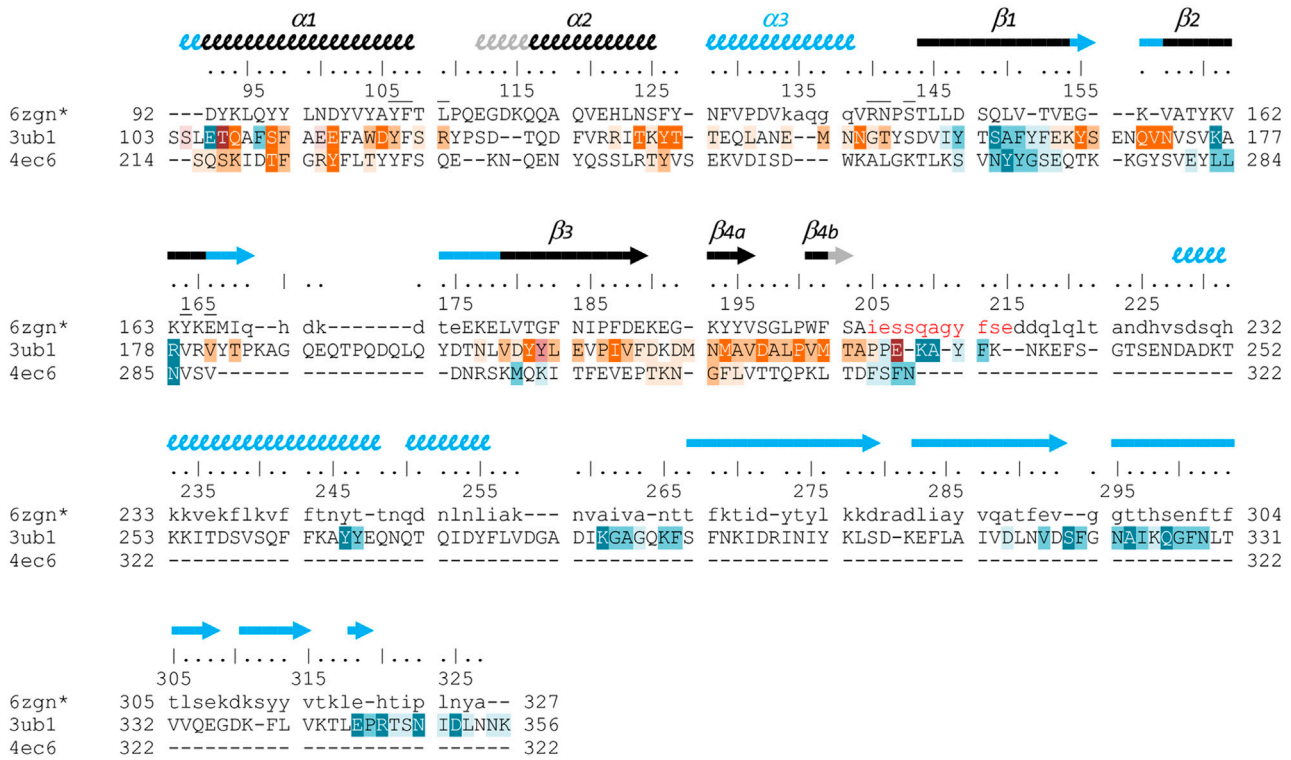


FIGURE 6 | Structure-based sequence alignment of three Gram-positive VirB8-like proteins. The sequence alignment was generated by mTM-align and manually modified, based on the three structures of $OrfG_{64–204}$ (pdb entry 6zgn), $TcpC_{99–359}$ (3ub1) and $TraM_{190–322}$ (4ec6). Residues of $OrfG_{64–204}$ that were not observed in the electron density are shown as lowercases, as well as residues 205–327 equivalent to the C-terminal domain of TcpC and for which no structure is known yet. Numbering above the sequences corresponds to OrfG. Secondary structures are represented by arrows (β-strands) and squiggles (α-helices), in black when shared by OrfG and TcpC, in grey when only present in OrfG, and in cyan when only present in TcpC. Residues of TraM and TcpC involved in the trimer assemblies are highlighted, in orange in one partner and in blue in the facing monomer: white letters with dark background are for residues involved in hydrogen bonds in all interface instances in the asymmetric units, while dark letters with lighter background are involved in contact in at least one interface (medium background when at least 50% of the residue is buried in the contact, light background if less than 50% is buried). Residues of $OrfG_{64–204}$ that delineate the pocket in which an unknown ligand is bound (see **Supplementary Figure S1**) are marked with an overbar.

their central domain, despite a low sequence identity, suggests a similar overall structure of the whole protein, and a similar oligomeric state. The missing linker and C-terminal domain probably hinder trimerization in the crystallization conditions used.

In Vitro Self-Oligomerization of $OrfG_{64–331}$ and Its Subdomains

The closest structural homologs to $OrfG_{64–204}$ form trimers in their crystal arrangement (**Figures 5B,C**). Thus, it was proposed that trimers constitute the functional biological assembly of both proteins (Porter et al., 2012, Goessweiner-Mohr et al., 2013b). These observations strongly call for the assessment of the oligomeric state of $OrfG_{64–204}$ in solution. Protein multimerization can be monitored by chemical cross-linking. We performed *in vitro* chemical cross-linking experiments with a constant amount of the protein in presence of an increasing concentration of the cross-linking agent paraformaldehyde (PFA). As observed in **Figure 7A**, in the presence of a low amount of PFA (0.05–1%), $OrfG_{64–331}$ predominantly forms dimers. High molecular-mass oligomers of $OrfG_{64–331}$ are visible at higher PFA concentrations (2 and 5%) demonstrating the predisposition of $OrfG_{64–331}$ to oligomerize in solution. The ability of $OrfG_{64–331}$ to self-assemble was confirmed by the analysis of the size exclusion chromatography (SEC) profile where the 31 kDa $OrfG_{64–331}$ was eluted into two defined peaks. SDS-PAGE analysis showed that both peaks are exclusively composed by $OrfG_{64–331}$ protein (**Figure 7B**). These results suggest that $OrfG_{64–331}$ adopts two oligomeric states in solution. To further examine the apparent existence of these two oligomeric states of $OrfG_{64–331}$, we analyzed purified $OrfG_{64–331}$ by SEC coupled to multiple angle light scattering (SEC-MALS) (**Figure 8A**). The predominant form (97.3% of the total mass fraction) was eluted at 29 min and corresponds to the monomeric form of $OrfG_{64–331}$ with a calculated MW of 32 kDa (theoretical MW = 31 kDa). The less present form (2.7% of the total mass fraction) was eluted earlier at 25 min with a calculated MW of 172 ± 11 kDa which can be assigned as 6-mer of $OrfG_{64–331}$ (186 kDa theoretical MW). SEC-

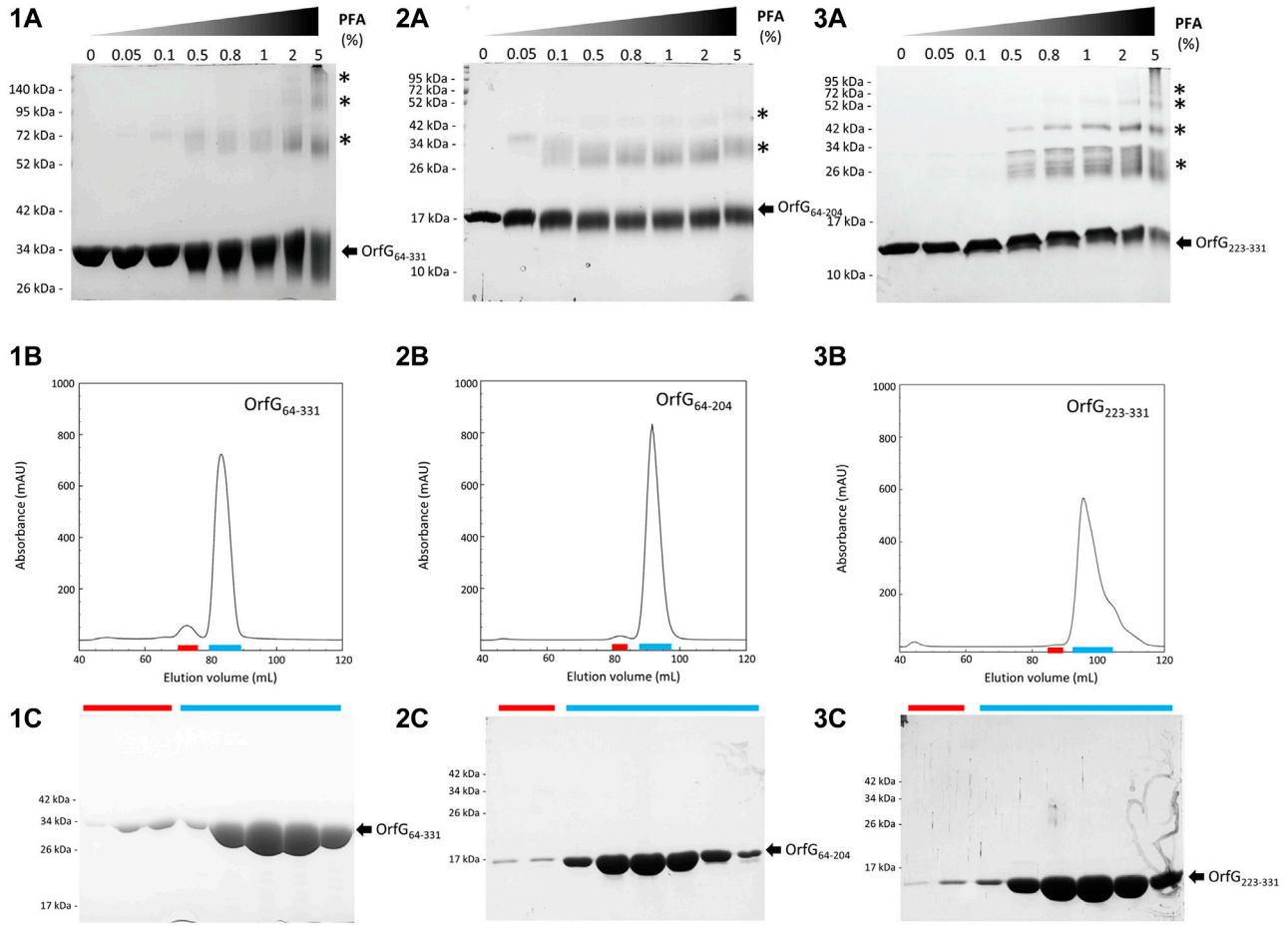


FIGURE 7 | Analysis of oligomerization of native and truncated versions of $OrfG_{64–331}$ by *in vitro* chemical cross-linking and size exclusion chromatography. *In vitro* chemical cross-linking of $OrfG_{64–331}$ (**1A**), $OrfG_{64–204}$ (**2A**) and $OrfG_{223–331}$ (**3A**). SDS-PAGE analysis of 3 μM of the purified $OrfG_{64–331}$ and its truncated versions in absence or in presence of an increasing concentration (0.05–5%) of paraformaldehyde (PFA). Samples were loaded without heating treatment. The concentration of PFA used for each reaction is mentioned at the top of each column. Asterisks indicate the identified oligomers stabilized by PFA cross-linking. Size exclusion chromatography (SEC) of the purified $OrfG_{64–331}$ (**1B**), $OrfG_{64–204}$ (**2B**) and $OrfG_{223–331}$ (**3B**) using Superdex 200 16/600. The elution volume (ml) is plotted on the *x*-axis and the 280-nm absorbance is plotted on the *y*-axis. SDS-PAGE analysis of SEC purification of $OrfG_{64–331}$ (**1C**), $OrfG_{64–204}$ (**2C**) and $OrfG_{223–331}$ (**3C**). Red and green lines correspond to the SEC fractions selected from each peak and analyzed by SDS-PAGE. Electrophoretic separation shows that all identified peaks contain exclusively the analyzed protein. Molecular weight markers (in kDa) are indicated on the left of each SDS-PAGE pattern. Protein bands corresponding to each protein are indicated by black arrows.

MALS data confirm the ability of $OrfG_{64–331}$ to form ordered homo-multimers in solution. Intriguingly, such behavior has not been reported for other characterized members of the Gram-positive VirB8-like proteins. In fact, AUC and SEC analysis performed for TcpC and TraH truncated for their TMD, respectively, revealed that both proteins were strictly monomeric in solution (Porter et al., 2012; Fercher et al., 2016). $TraM_{190–322}$ (TraM lacking its N-terminal domain and the TMD) was able to form oligomers only in the presence of cross-linking agents whereas DLS and SAXS analysis indicated a monomeric form of TraM in solution (Goessweiner-Mohr et al., 2013).

The knowledge that OrfG can assemble into ordered oligomers in solution led us to question whether OrfG central domain or OrfG C-terminal domain or both would promote/contribute to $OrfG_{64–331}$ oligomerization. For that purpose, we investigate separately their capacity to self-assemble by using chemical cross-linking. Regarding $OrfG_{64–204}$, the *in vitro* chemical cross-linking using PFA gave mostly a dimeric form when using at least 0.05% of PFA (**Figure 7C**). At higher PFA concentrations, an additional band with a higher mass was observed that might correspond to a trimeric form. On the contrary to $OrfG_{64–204}$, the chemical cross-linking of $OrfG_{223–331}$ leads to the formation of multiple oligomers of $OrfG_{223–331}$ easily distinguished when analyzed by SDS-PAGE (**Figure 7E**). These results suggest that both subdomains were independently able to form oligomers in solution. Interestingly, during SEC analysis of both $OrfG_{64–204}$ and $OrfG_{223–331}$, we observed the presence of two well-defined peaks as

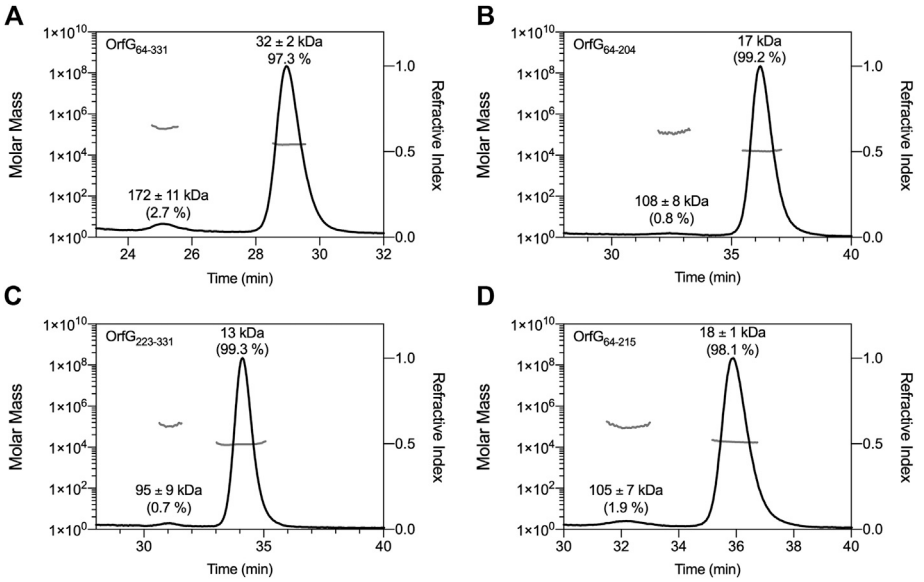


FIGURE 8 | SEC-MALS analysis. Elution profile (black lines) of $OrfG_{64–331}$ **(A)**, $OrfG_{64–204}$ **(B)**, $OrfG_{223–331}$ **(C)** and $OrfG_{64–215}$ **(D)** are shown with the molecular weight calculated by MALS (gray lines). The elution time (min) is plotted on the *x*-axis. The molar mass (in logarithmic scale) is plotted in the first *y*-axis and the refractive index is plotted in the second *y*-axis. The molecular weight (in kDa) and the contribution in mass fraction (in %) of each visible peak are shown at the top of the corresponding peak.

seen on the SEC profile of $OrfG_{64–331}$ (**Figures 7B,D,F**). These findings prompted us to characterize further the oligomerization of $OrfG_{64–204}$ and $OrfG_{223–331}$ using SEC-MALS. As observed in **Figures 8B,C**, both $OrfG_{64–204}$ and $OrfG_{223–331}$ form two oligomeric states. A predominant monomeric form was eluted at 36 min for $OrfG_{64–204}$ and at 34 min for $OrfG_{223–331}$ with a calculated molar mass of 17 and 13 kDa, respectively. Additional early peaks were barely detected at 32.5 min and 31 min with calculated molar mass of 108 ± 8 and 95 ± 9 kDa for $OrfG_{64–204}$ and $OrfG_{223–331}$, respectively. Interestingly, calculated molar masses for the ordered oligomers detected for the $OrfG_{64–331}$ subdomains were compatible with 6-mer assembly for each subdomain (102 and 78 kDa theoretical MW for $OrfG_{64–204}$ and $OrfG_{223–331}$, respectively).

DISCUSSION

During the last decades, a great effort was made to understand the mode of action of conjugative type IV secretion systems (Conj-T4SSs) due to their major role in antibiotic resistance spreading among bacteria. More attention was devoted to the study of Conj-T4SS from Gram-negative bacteria. Multiple structural information has been gained for individual components/proteins and isolated complexes improving our understanding of their architecture and assembly mode (reviewed in (Grohmann et al., 2018)). Conj-T4SS from Gram-positive bacteria have been less investigated resulting in little knowledge on their architecture and on the molecular mechanisms underlying DNA transfer in these bacteria (reviewed in (Goessweiner-Mohr et al., 2013a)).

In order to gain further information on the architecture and the assembly mode of Conj-T4SS from Gram-positive bacteria, we focused on the structural, biochemical and biophysical study of OrfG, a putative transfer protein encoded within the conjugative module of ICE*St3* from *S. thermophilus*. By using X-ray crystallography, we solved the structure of the central domain of OrfG. Structural analysis revealed that OrfG central domain adopts an NTF2-like fold, common to all known VirB8-like proteins in both Gram-negative and -positive Conj-T4SSs. VirB8-like proteins are essential structural and functional component of Conj-T4SSs. Consequently, we propose that OrfG operates as a VirB8-like protein in the conjugative apparatus of ICE*St3*. The structure of OrfG central domain represents the first structure of a VirB8-like protein from ICEs described so far and rises to four the VirB8-like protein solved structures for Gram-positive Conj-T4SS. Notwithstanding the tricky task to identify VirB8-like proteins in Gram-positive Conj-T4SS clusters due to their low sequence identity, the growing number of VirB8-like structures indicates that it is a conserved protein in conjugation machines and underlines the similarity between protein complexes that govern the transport of mobile genetic elements across the cytoplasmic membrane in both Gram-negative and -positive bacteria.

OrfG central domain shares strong resemblance with the TcpC central domain of pCW3 from *C. perfringens*. Despite their low sequence similarity/identity, *in silico* analysis (*see Materials and Methods* section) suggests that both OrfG and TcpC adopt the same domain organization with an N-terminal transmembrane domain (TMD) followed by two soluble subdomains. This domain organization appears to be conserved in a large number of putative transfer proteins from both conjugative

plasmids and ICEs alike TcpC from pJIR26, Orf13 from ICECW459, Tn*916* and Tn*5397* and the VirB8-like protein (conB) from ICE*Bs1*. Our phylogenetic analysis supports this observation showing that these proteins belong to the TcpC family (**Figure 1B**).

In addition to TcpC and OrfG, two other transfer proteins, TraM and TraH from the conjugative plasmid pIP501 from *E. faecalis,* hold a NTF2-like domain in their C-terminal domains and were proposed to be VirB8-like proteins. Interestingly, both TraM and TraH adopt a different domain organization compared to TcpC and OrfG (**Figure 1A**). In fact, TraM is composed of two soluble subdomains separated by a TMD while TraH is composed of an N-terminal TMD followed by a unique C-terminal subdomain. The variability in domain organization is restricted to VirB8-like proteins from Gram-positive bacteria since VirB8-like proteins from Gram-negative bacteria share a common domain organization with an N-terminal TMD followed by a single NTF2-like C-terminal periplasmic domain. The modularity/flexibility in domain organization perceived for Gram-positive VirB8-like proteins could be explained in part by the evolutionary scenario of conjugative systems. In fact, it was proposed that Gram-positive conjugative systems emerge from Gram-negative ones by gene deletion and such transfer was followed by diverse adaptive routes (Guglielmini et al., 2013). This significant variation could have originated from the varied cell envelope composition and thickness and/or the functional adaptation of conjugative systems regarding the nature of the donor and recipient cells as well as the variation of the conjugation environment (Alvarez-Martinez and Christie, 2009). Actually, most of the NTF2-like domains of the VirB8-like proteins from Gram-positive Conj-T4SS are localized in the cell-wall, such localization makes these domains more exposed to the environmental stress and this may contribute to their modularity compared to the periplasmic localization of VirB8-like protein in Gram-negative bacteria (Christie, 2016; Waksman, 2019).

Strikingly, although they are subtle, the differences observed when comparing the structures of all the NTF2-like domains of VirB8-like proteins lead again to distinguish Gram-positive and -negative bacteria. The tree based on structural similarities seems significant as it groups structures in a way comparable to the one obtained through phylogenetic analysis (when applicable i.e., for sequences that belong to the PFAM group PF04335, *see* **Figure 4**). Moreover, this tree, strictly based on the spatial superposition of atoms without valuing sequence conservation, associates in a same class, proteins that share low sequence similarity. For instance, in the case of Gram-negative bacteria with more known structures, VirB8s from *A. tumefaciens* and from *Bartonella quintana* gather in class II^{-} although they only have 22% of sequence identity; Class III^{-} contains TraM from Plasmid R64 (Kuroda et al., 2015) and the IcmL-like protein from *Legionella pneumophila* despite only 13% of identity (**Supplementary Table S2**). Thus, the partition of VirB8-like structures from Gram-negative bacteria on one side, and Gram-positive ones on the other side in the tree is highly significant since the low sequence similarity of the latter with any other could have distribute them anywhere in the tree. However, if clear in the tree, this distinction is less apparent when looking at the structures themselves, so that it is hard to define peculiar features that cause this distinction. Enrichment of the database by new structures of VirB8-like proteins from Gram-positive bacteria will probably bring elements that underpin this partition, as well as it will certainly draw several new classes in this part of the tree. The current distribution separates the monomeric *E. faecalis* TraH/Orf8 from the class I^{+} which contains TcpC from *C. perfringens* and TraM from *E. faecalis*. The presence of OrfG with these two members that share the same trimeric crystal packing questions the physiological assembly of this protein.

One of the most puzzling questions on the study of bacterial secretion systems including the Conj-T4SS concerns the oligomeric state and dynamics of their components inside the cells. In Conj-T4SS, VirB8 proteins are bitopic proteins of the cytoplasmic membrane. In Gram-negative bacteria, many studies suggested/supported that VirB8 function as a multimers. Most of the periplasmic domains of all studied VirB8 proteins adopt a dimeric assembly in solution and it is believed that these dimers are the forming subunits of higher-order multimers which should be structural parts of the functional Conj-T4SS (Casu et al., 2018). Structural and biochemical data proposed that the full-length VirB8 protein self-assembles into a homo-multimers with C3 symmetry. Indeed, the stoichiometry of VirB8 was estimated to 12 in Conj-T4SS from *E. coli* R388 (Low et al., 2014). In line with, the purification of full-length TraE and TraM, a VirB8 homolog from pKM101 and R64, respectively, revealed that TraE and TraM forms hexamers in solution (Kuroda et al., 2015; Casu et al., 2018). Dissimilar to what was described in Gram-negative bacteria, VirB8-like proteins in Gram-positive bacteria are proposed to function as trimers *in vivo* (Porter et al., 2012, Goessweiner-Mohr et al., 2013). Up to date, no higher-order oligomers were obtained for VirB8-like proteins from Gram-positive bacteria supporting a different assembly mode compared to VirB8-like proteins recovered in Gram-negative bacteria. Nevertheless, this variation in the assembly mode do not amend the capacity of VirB8-like proteins to self-assemble following a C3 symmetry independently from the origin of the conjugative system. For Gram-positive VirB8-like proteins, TMD was proposed to trigger trimeric assembly since isolated TMD-less VirB8-like sub-domains were monomeric in solution. In addition, a predicted coiled-coil motif preceding the NTF2-like domain, not conserved in all VirB8-like proteins, seems to play an important role in the stabilization of the trimeric assembly (Goessweiner-Mohr et al., 2013b). In this context, our in-solution characterization revealed intriguing properties concerning OrfG: 1) significant but small fraction of OrfG soluble domain multimerize independently of its TMD and/or a coiled-coil domain since *in silico* analyses failed to find a predicted coiled-coil motif in the purified OrfG soluble domain, 2) OrfG soluble domain has the ability to spontaneously assembles into 6-mer. This ordered oligomeric state observed in solution appears to be a physiological property of OrfG since SEC-MALS analysis performed on OrfG central and C-terminal sub-domains revealed that both form the same ordered-oligomers in solution. The low proportion of OrfG oligomers observed in solution could be explained by the absence of the TMD known to be essential for VirB8

oligomers stabilization (Kuroda et al., 2015; Casu et al., 2018). These data exposed a discrepancy on the number of the forming subunits between OrfG 6-mer oligomers and TcpC and TraM trimers. This could be explained by the fact that OrfG 6-mer oligomers could result from the dimerization of trimeric complexes. Another probable explanation lies on the ability of OrfG to adopt a different assembly mode not yet described for Gram-positive VirB8-like proteins.

Nevertheless, our results combined with published data obtained for TcpC, TraM and TraH consolidate the fact that VirB8-like proteins from Gram-positive bacteria act as multimeric entity on their biological systems but interrogates on their physiological stoichiometry that may depends on their dedicated systems.

VirB8 proteins from Gram-negative bacteria were considered to be interesting targets to develop specific inhibitors of Conj-T4SS by abolishing VirB8 dimerization. These inhibitors interact with a specific site localized in a structurally conserved groove not directly involved in the dimerization of VirB8 proteins (Smith et al., 2012; Casu et al., 2016). Since the inhibitor-binding interface is on the opposite site of dimerization interface, it was proposed that conformational changes on this groove directly affect VirB8 dimerization. Surprisingly, the electron density of OrfG central domain displayed an unassigned blob (**Supplementary Figure S1**) nicely docked in a pocket localized in a hydrophobic cavity that shares intriguing similarity with the groove identified in Gram-negative VirB8 proteins. In fact, residues located on the C-terminal end of α_1 and β_4 delimit both of the identified pockets. Thus, the localization of this groove suggests a probable oligomerization inhibition site for OrfG. The presence of this pocket in VirB8-like proteins could be explained by the nature of the NTF2-like fold. In fact, this fold is widely distributed in bacteria and is associated with various functions including enzymatically active and non-enzymatically active proteins. When extracellular, these proteins often possess non-catalytic ligand-binding activities (Eberhardt et al., 2013).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTION

JC, AM-A, SM, TD, CD, and BD performed the experiments. NL-B phylogeny analysis. JC developed the script used for structural analysis. CD, FF, NL-B, and BD supervision, validation, investigation, methodology. FF and BD Writing original draft. All authors: review and editing of the manuscript.

FUNDING

This work was funded by the National Research Institute for Agriculture, Food and Environment INRAE, the Centre National de la Recherche Scientifique CNRS, the Université de Lorraine and its A2F and CPM scientific poles.

ACKNOWLEDGMENTS

We thank the members of the DynAMic and CRM2 laboratories for insightful discussions; Laurence Hotel, Emilie Piotrowski, Louise Thiriet, Stéphane Bertin, Johan Staub and Anthony Gauthier for technical assistance; Nicolas Soler and Yvonne Roussel for valuable discussions; Emilie Robert and Jean-Michel Girardet from ASIA platform for encouragements. SEC-MALS analysis were conducted using MiniDAWN TREOS II mounted on FPLC system (AKTA purifier) available on the ASIA platform (Université de Lorraine-INRAE, https://a2f.univ-lorraine.fr/asia/). The Soleil Synchrotron radiation facility is acknowledged for beamline allocation. The authors appreciated the access to the 'Plateforme de mesures de diffraction X' of the Université de Lorraine.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.642606/full#supplementary-material.

REFERENCES

Alvarez-Martinez, C. E., and Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* 73, 775–808. doi:10.1128/MMBR.00023-09

Ambroset, C., Coluzzi, C., Guédon, G., Devignes, M. D., Loux, V., Lacroix, T., et al. (2015). New insights into the classification and integration specificity of Streptococcus integrative conjugative elements through extensive genome exploration. *Front. Microbiol.* 6, 1483. doi:10.3389/fmicb.2015.01483

Arends, K., Celik, E. K., Probst, I., Goessweiner-Mohr, N., Fercher, C., Grumet, L., et al. (2013). TraG encoded by the pIP501 type IV secretion system is a two-domain peptidoglycan-degrading enzyme essential for conjugative transfer. *J. Bacteriol.* 195, 4436–4444. doi:10.1128/JB.02263-12

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. doi:10.1093/nar/gkw408

Bantwal, R., Bannam, T. L., Porter, C. J., Quinsey, N. S., Lyras, D., Adams, V., et al. (2012). The peptidoglycan hydrolase TcpG is required for efficient conjugative transfer of pCW3 in *Clostridium perfringens*. *Plasmid* 67, 139–147. doi:10.1016/j.plasmid.2011.12.016

Bellanger, X., Roberts, A. P., Morel, C., Choulet, F., Pavlovic, G., Mullany, P., et al. (2009). Conjugative transfer of the integrative conjugative elements ICESt1 and ICESt3 from *Streptococcus thermophilus*. *J. Bacteriol.* 191, 2764–2775. doi:10.1128/JB.01412-08

Bhatty, M., Laverde Gomez, J. A., and Christie, P. J. (2013). The expanding bacterial type IV secretion lexicon. *Res. Microbiol.* 164, 620–639. doi:10.1016/j.resmic.2013.03.012

Buchan, D. W. A., and Jones, D. T. (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* 47, W402–W407. doi:10.1093/nar/gkz297

Burrus, V., Pavlovic, G., Decaris, B., and Guédon, G. (2002). The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* 48, 77–97. doi:10.1016/s0147-619x(02)00102-6

Cascales, E., and Christie, P. J. (2003). The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* 1, 137–149. doi:10.1038/nrmicro753

Casu, B., Mary, C., Sverzhinsky, A., Fouillen, A., Nanci, A., and Baron, C. (2018). VirB8 homolog TraE from plasmid pKM101 forms a hexameric ring structure and interacts with the VirB6 homolog TraD. *Proc. Natl. Acad. Sci. USA* 115, 5950–5955. doi:10.1073/pnas.1802501115

Casu, B., Smart, J., Hancock, M. A., Smith, M., Sygusch, J., and Baron, C. (2016). Structural analysis and inhibition of TraE from the pKM101 type IV secretion system. *J. Biol. Chem.* 291, 23817–23829. doi:10.1074/jbc.M116.753327

Christie, P. J. (2016). The mosaic type IV secretion systems. *EcoSal Plus* 7. doi:10.1128/ecosalplus.ESP-0020-2015

Dahmane, N., Robert, E., Deschamps, J., Meylheuc, T., Delorme, C., Briandet, R., et al. (2018). Impact of cell surface molecules on conjugative transfer of the integrative and conjugative element ICESt3 of *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* 84, e02109. doi:10.1128/AEM.02109-17

Dong, R., Pan, S., Peng, Z., Zhang, Y., and Yang, J. (2018a). mTM-align: a server for fast protein structure database search and multiple protein structure alignment. *Nucleic Acids Res.* 46, W380–W386. doi:10.1093/nar/gky430

Dong, R., Peng, Z., Zhang, Y., and Yang, J. (2018b). mTM-align: an algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* 34, 1719–1725. doi:10.1093/bioinformatics/btx828

Dunny, G. M. (2007). The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 362, 1185–1193. doi:10.1098/rstb.2007.2043

Eberhardt, R. Y., Chang, Y., Bateman, A., Murzin, A. G., Axelrod, H. L., Hwang, W. C., et al. (2013). Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics* 14, 327. doi:10.1186/1471-2105-14-327

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi:10.1107/S0907444910007493

Evans, P. R., and Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* 69, 1204–1214. doi:10.1107/S0907444913000061

Fercher, C., Probst, I., Kohler, V., Goessweiner-Mohr, N., Arends, K., Grohmann, E., et al. (2016). VirB8-like protein TraH is crucial for DNA transfer in *Enterococcus faecalis*. *Sci. Rep.* 6, 24643. doi:10.1038/srep24643

Fronzes, R., Christie, P. J., and Waksman, G. (2009). The structural biology of type IV secretion systems. *Nat. Rev. Microbiol.* 7, 703–714. doi:10.1038/nrmicro2218

Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732. doi:10.1038/nrmicro1235

Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., et al. (2020). Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinformatics* 72, e108. doi:10.1002/cpbi.108

Goessweiner-Mohr, N., Arends, K., Keller, W., and Grohmann, E. (2013a). Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid* 70, 289–302. doi:10.1016/j.plasmid.2013.09.005

Goessweiner-Mohr, N., Grumet, L., Arends, K., Pavkov-Keller, T., Gruber, C. C., Gruber, K., et al. (2013b). The 2.5 Å structure of the enterococcus conjugation protein TraM resembles VirB8 type IV secretion proteins. *J. Biol. Chem.* 288, 2018–2028. doi:10.1074/jbc.M112.428847

González-Rivera, C., Khara, P., Awad, D., Patel, R., Li, Y. G., Bogisch, M., et al. (2019). Two pKM101-encoded proteins, the pilus-tip protein TraC and Pep, assemble on the *Escherichia coli* cell surface as adhesins required for efficient conjugative DNA transfer. *Mol. Microbiol.* 111, 96–117. doi:10.1111/mmi.14141

Grohmann, E., Christie, P. J., Waksman, G., and Backert, S. (2018). Type IV secretion in gram-negative and gram-positive bacteria. *Mol. Microbiol.* 107, 455–471. doi:10.1111/mmi.13896

Guglielmini, J., de la Cruz, F., and Rocha, E. P. (2013). Evolution of conjugation and type IV secretion systems. *Mol. Biol. Evol.* 30, 315–331. doi:10.1093/molbev/mss221

Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. (2011). The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *Plos Genet.* 7, e1002222. doi:10.1371/journal.pgen.1002222

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi:10.1093/sysbio/syq010

Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* 20, 478–480. doi:10.1016/s0968-0004(00)89105-7

Ilangovan, A., Kay, C. W. M., Roier, S., El Mkami, H., Salvadori, E., Zechner, E. L., et al. (2017). Cryo-EM structure of a relaxase reveals the molecular basis of DNA unwinding during bacterial conjugation. *Cell* 169, 708–721. doi:10.1016/j.cell.2017.04.010

Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132. doi:10.1107/S0907444909047337

Kohler, V., Keller, W., and Grohmann, E. (2018). Enterococcus adhesin PrgB facilitates type IV secretion by condensation of extracellular DNA. *Mol. Microbiol.* 109, 263–267. doi:10.1111/mmi.13994

Kohler, V., Probst, I., Aufschnaiter, A., Büttner, S., Schaden, L., Rechberger, G. N., et al. (2017). Conjugative type IV secretion in Gram-positive pathogens: TraG, a lytic transglycosylase and endopeptidase, interacts with translocation channel protein TraM. *Plasmid* 91, 9–18. doi:10.1016/j.plasmid.2017.02.002

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi:10.1016/j.jmb.2007.05.022

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256–2268. doi:10.1107/S0907444904026460

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Kuroda, T., Kubori, T., Thanh Bui, X., Hyakutake, A., Uchida, Y., Imada, K., et al. (2015). Molecular and structural analysis of Legionella DotI gives insights into an inner membrane complex essential for type IV secretion. *Sci. Rep.* 5, 10912. doi:10.1038/srep10912

Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* 3, 1171–1179. doi:10.1038/nprot.2008.91

Laverde Gomez, J. A., Bhatty, M., and Christie, P. J. (2014). PrgK, a multidomain peptidoglycan hydrolase, is essential for conjugative transfer of the pheromone-responsive plasmid pCF10. *J. Bacteriol.* 196, 527–539. doi:10.1128/JB.00950-13

Leonetti, C. T., Hamada, M. A., Laurer, S. J., Broulidakis, M. P., Swerdlow, K. J., Lee, C. A., et al. (2015). Critical components of the conjugation machinery of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *J. Bacteriol.* 197, 2558–2567. doi:10.1128/JB.00142-15

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* 75, 861–877. doi:10.1107/S2059798319011471

Low, H. H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., et al. (2014). Structure of a type IV secretion system. *Nature* 508, 550–553. doi:10.1038/nature13081

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162–1164. doi:10.1126/science.252.5009.1162

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674. doi:10.1107/S0021889807021206

Porter, C. J., Bantwal, R., Bannam, T. L., Rosado, C. J., Pearce, M. C., Adams, V., et al. (2012). The conjugation protein TcpC from *Clostridium perfringens* is structurally related to the type IV secretion system protein VirB8 from Gram-negative bacteria. *Mol. Microbiol.* 83, 275–288. doi:10.1111/j.1365-2958.2011.07930.x

Roberts, A. P., and Mullany, P. (2011). Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol. Rev.* 35, 856–871. doi:10.1111/j.1574-6976.2011.00283.x

Santoro, F., Vianna, M. E., and Roberts, A. P. (2014). Variation on a theme; an overview of the Tn916/Tn1545 family of mobile genetic elements in the oral and nasopharyngeal streptococci. *Front. Microbiol.* 5, 535. doi:10.3389/fmicb.2014.00535

Schroeder, M. R., and Stephens, D. S. (2016). Macrolide resistance in *Streptococcus pneumoniae*. *Front Cell Infect Microbiol* 6, 98. doi:10.3389/fcimb.2016.00098

Smith, M. A., Coinçon, M., Paschos, A., Jolicoeur, B., Lavallée, P., Sygusch, J., et al. (2012). Identification of the binding site of Brucella VirB8 interaction inhibitors. *Chem. Biol.* 19, 1041–1048. doi:10.1016/j.chembiol.2012.07.007

Soler, N., Robert, E., Chauvot de Beauchêne, I., Monteiro, P., Libante, V., Maigret, B., et al. (2019). Characterization of a relaxase belonging to the MOBT family, a widespread family in Firmicutes mediating the transfer of ICEs. *Mob DNA* 10, 18. doi:10.1186/s13100-019-0160-9

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526. doi:10.1093/oxfordjournals.molbev.a040023

Waksman, G. (2019). From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. *EMBO Rep.* 20. doi:10.15252/embr.201847012

Wisniewski, J. A., and Rood, J. I. (2017). The Tcp conjugation system of *Clostridium perfringens*. *Plasmid* 91, 28–36. doi:10.1016/j.plasmid.2017.03.001

Wisniewski, J. A., Teng, W. L., Bannam, T. L., and Rood, J. I. (2015). Two novel membrane proteins, TcpD and TcpE, are essential for conjugative transfer of pCW3 in *Clostridium perfringens*. *J. Bacteriol.* 197, 774–781. doi:10.1128/JB.02466-14

Wolf, E., Kim, P. S., and Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.* 6, 1179–1189. doi:10.1002/pro.5560060606

Wu, X., Zhao, Y., Sun, L., Jiang, M., Wang, Q., Wang, Q., et al. (2019). Crystal structure of CagV, the *Helicobacter pylori* homologue of the T4SS protein VirB8. *FEBS J.* 286, 4294–4309. doi:10.1111/febs.14971

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430, 2237–2243. doi:10.1016/j.jmb.2017.12.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cappele, Mohamad Ali, Leblond-Bourget, Mathiot, Dhalleine, Payot, Savko, Didierjean, Favier and Douzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.