Supplementary information for:

The outer-membrane export signal of *Porphyromonas gingivalis* type IX secretion system (T9SS) is a conserved C-terminal β-sandwich domain

Iñaki de Diego Martinez^{1,§,#}, Miroslaw Ksiazek^{2,3,4,§}, Danuta Mizgalska², Lahari Koneru⁴, Przemyslaw Golik², Borys Szmigielski², Magdalena Nowak², Zuzanna Nowakowska², Barbara Potempa⁴, John A. Houston⁴, Jan J. Enghild⁵, Ida B. Thøgersen⁵, Jinlong Gao^{6,7}, Ann H. Kwan⁸, Jill Trewhella⁸, Grzegorz Dubin^{2,3}, F. Xavier Gomis-Rüth¹, Ky-Anh Nguyen^{6,7,*} and Jan Potempa^{2,3,4,*}

Table S1. X-ray data collection and refinement statistics

Protein name	rCTD	r664i6H	r665sXa	r665i6H -	
PDB entry	5HFS	5AG8	5AG9		
		X-ray data collection			
Space group	P65	P21	P2 ₁	P2 ₁	
Wavelength (Å)	1.5419	0.9724	0.8726	0.8726	
Unit cell constants <i>a</i> , <i>b</i> , <i>c</i> (Å), â (°)	37.69, 37.69, 184.15, 90.0	51.47, 61.74, 55.30, 117.8	51.6, 61.8, 54.1, 116.8	52.9, 63.2, 56.7, 117.8	
Resolution (Å) ^a	32.7-1.97 (2.17-1.97)	48.9-1.90(2.0-1.90)	38.0-2.11(2.22-2.11)	39.3-2.44(2.57-2.44)	
Completeness (%)	100.0 (100.0)	98.6 (91.9)	99.9 (100.0)	95.3 (95.5)	
No. of measurements / Unique reflections	58,366 / 10,507	86,628 / 23,941	65,902 /17,553	28,402 / 11,842	
R _{merge}	0.121 (0.411)	0.049 (0.270)	0.150 (0.447)	0.158 (0.447)	
Average intensity (<i> / s(I))</i>	7.0 (2.7)	20.5 (4.5)	6.4 (2.9)	4.9 (2.2)	
Redundancy	5.6 (5.3)	3.6 (3.3)	3.8 (4.0)	2.4 (2.4)	
		Refinement			
Resolution (Å)	∞-1.97	∞-1.90	∞-2.10		
No. reflections	No. reflections 10,402		16,650		
$\mathbf{R}_{\mathrm{cryst}}$ / $\mathbf{R}_{\mathrm{free}}$	R _{cryst} / R _{free} 0.20/0.24		0.19/0.25		

No. atoms (protein / ligands / solvent)	atoms (protein / ligands / 944/6/33 solvent)		2208/25/254	
Thermal-displacement parameters (Å ² ; protein / ligands / solvent)	39.3/93.2/40.0	23.1/49.9/33.1	19.3/53.5/28.6	
R.m.s.d. bond lengths (Å)	0.007	0.019	0.016	
R.m.s.d. bond angles (°)	1.12	1.98	1.92	
Residues in favoured/allowed regions (%) ^b	95.2/100.0	97.2/99.7	97.6/99.7	

^a Values in parentheses refer to highest resolution shell.

^b According to MOLPROBITY (Davis et al., 2007, Chen et al., 2010).

 R_{free} is the same as R_{value} except that calculated for a set of the data (>500) randomly omitted from refinement.

Chen et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. sect. D 66, 12-21 (2010).

Davis *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucl. Acids Res.* **35**, W375-W383 (2007).

Evans. "Scaling and assessment of data quality." Acta Crystallogr. sect. D 62, 72-82 (2006).

García-Castellanos *et al.* Three-dimensional structure of MecI : Molecular basis for transcriptional regulation of staphylococcal methicillin resistance. *J. Biol. Chem.* **278**, 39897-39905 (2003).

Kabsch. XDS. Acta Crystallogr. sect. D 66, 125-132 (2010).

Sztukowska *et al.* The C-terminal domains of the gingipain K polyprotein are necessary for assembly of the active enzyme and expression of associated activities. *Mol. Microbiol.* **54**, 1393-1408 (2004).

Weiss. Global indicators of X-ray quality. J. Appl. Cryst. 34, 130-135 (2001).

	rlgSF-CTD (wild-type)	r664i6H		
Guinier analysis*				
$R_{g}(m \AA)$	24 ± 1	25 ± 1		
<i>qR_g</i> -range	0.5 – 1.18	0.5 - 1.15		
Fidelity	0.99	0.98		
P(r) analysis*				
$R_{g}(\text{\AA})$	25.2 ± 0.1	26.0 ± 0.2		
d _{max} (Å)	98	97		
<i>q</i> -range (Å ⁻¹)	0.02-0.23	0.02-0.23		
total estimate	0.68	0.67		
MW from I(0) (Da)	16619	15012		
Sample details				
Concentration range measured (mg mL ⁻¹)	0.2 - 1.7	0.14 – 1.4		
MW from sequence (Da)	17030	17849		
Ratio of <i>MW</i> from <i>I</i> (0) to <i>MW</i> from sequence	0.98	0.84		
Ratio of <i>MW</i> from DAMMIF to <i>MW</i> from sequence	0.93	0.94		
Partial specific volume from sequence (v , cm ³ g ⁻¹)	0.741	0.739		
Contrast from sequence and solvent constituents ($\Delta \rho$, 10^{10} cm^{-2})	2.790	2.814		
Model fitting results				
DAMMIF NSD (std dev)	0.684 (0.031)	0.654 (0.042)		
DAMMIF χ^2 range	1.123 - 1.146	1.031 – 1.041		
DAMMIF MW (Da)	15800	16700		
BUNCH χ ²	1.312	1.287		
EOM χ^2	1.04	0.97		
* Parameters reported for averaged $I(q)$ vs q of highest concentration measurement for two repeated dilution series; 10 – 100% dilutions. 100% conc. estimates 1.66 and 1.35 mg/mL for rlgSF-CTD and r664i6H, respectively.				

Table S2: SAXS Data Summary

Data collection parameters

Instrumentation	Australian Synchrotron SAXS beam-line
Beam geometry	544 μm x 274 μm

<i>q</i> -range measured (Å ⁻¹)	0.006-0.33			
Exposure time (seconds) 1 s x 24 frames				
Temperature (°C)	13.5			
Software employed for data reduction, analysis and interpretation				
SAXS data reduction	ScatterBrain			
Calculation of expected MW, $\Delta \rho$ and	MULCh			
SAXS data analysis: extrapolation to Guinier and $P(r)$	ATSAS 2.6.0 SAS Data Analysis			
ab initio dummy residue modelling	DAMMIF (via ATSAS on-line)			
Atomic structure modelling	BUNCH			
Ensemble modelling	EOM (via ATSAS on-line)			
3D graphic model representations	PYMOL			

Primer name	Sequence (5'-3')				
Recombinant protein cloning primers					
RgpBlgBamHIF	TAGGATCCACACTTGTCCCGACCAAAATGC				
RgpBCterBamHIF	TAGGATCCACATCTATTGCCGACGTAGCC				
RgpBCterXholR	AACTCGAGAACAGTCTCTTGGCGTAGTGC				
Thrombin cleavage	e to PreScission cleavage mutagenesis primers				
RThr2PreFA	CTGGAAGTTCTGTTCCAGGGTCCGACACTTGTCCCGACCAAAATG				
RThr2PreFB	ACACTTGTCCCGACCAAAATG				
GThr2PreRA	CGGACCCTGGAACAGAACTTCCAGATCGCTTTTTGGAGGATGGTCGCC				
GThr2PreRB	ATCGCTTTTTGGAGGATGGTCGCC				
662iXa mutagenes	is primers				
662iXaFl	ATCGAAGGCCGTGCAGCTACATCTATTGCCGACGTAGCC				
662iXaFs	ACATCTATTGCCGACGTAGCC				
662iXaRl	AGCTGCACGGCCTTCGATACCTTCCACTTTCACATCCTTTATCAC				
662iXaRs	ACCTTCCACTTTCACATCCTTTATCAC				
662i6H mutagenes	is primers				
662i6HFI	CATCACCATCACCATCACATCTATTGCCGACGTAGCC				
662iXaFs	ACATCTATTGCCGACGTAGCC				
662i6HRI	GTGATGGTGATGGTGATGACCTTCCACTTTCACATCCTTTATCAC				
662iXaRs	ACCTTCCACTTTCACATCCTTTATCAC				
662sXa mutagenes	sis primers				
662sXaFl	ATCGAAGGCCGTGCAGCTGTAGCCAATGATAAGCCTTAT				
662sXaFs	GTAGCCAATGATAAGCCTTAT				
662sXaRl	AGCTGCACGGCCTTCGATTTCCACTTTCACATCCTTTATC				
662sXaRs	TTCCACTTTCACATCCTTTATC				
664iXa mutagenes	is primers				
664iXaFl	ATCGAAGGCCGTGCAGCTATTGCCGACGTAGCCAATGATAAGC				
664iXaFs	ATTGCCGACGTAGCCAATGATAAGC				
664iXaRl	AGCTGCACGGCCTTCGATAGATGTACCTTCCACTTTCACATC				
664iXaRs	AGATGTACCTTCCACTTTCACATCCT				
664i6H mutagenes	is primers				
664i6HFI	CATCACCATCACCATCACATTGCCGACGTAGCCAATGATAAGC				
664iXaFs	ATTGCCGACGTAGCCAATGATAAGC				
664i6HRI	GTGATGGTGATGGTGATGAGATGTACCTTCCACTTTCACATC				
664iXaRs	AGATGTACCTTCCACTTTCACATCCT				
665i6H mutagenes	is primers				
665i6HFI	CATCACCATCACCATCACGCCGACGTAGCCAATGATA				
665i6HFs	GCCGACGTAGCCAATGATA				
665i6HRI	GTGATGGTGATGGTGATGAATAGATGTACCTTCCACTTTCAC				
665i6HRs	AATAGATGTACCTTCCACTTTCAC				
665sXa mutagenes	sis primers				
665sXaFI	ATCGAAGGCCGTGCAGCTGATAAGCCTTATACTGTAGCTGTATCAGG				
665sXaFs	GATAAGCCTTATACTGTAGCTGTATCAGG				
665sXaRI	AGCTGCACGGCCTTCGATAGATGTACCTTCCACTTTCACATCC				
665sXaRs	AGATGTACCTTCCACTTCACATCC				
665s6H mutagenes	șis primers				
665s6HFI	CATCACCATCACCATCACGATAAGCCTTATACTGTAGCTGT				
665s6HFs	GATAAGCCTTATACTGTAGCTGT				
665s6HRI	GTGATGGTGATGGTGATGAGATGTACCTTCCACTTTCAC				
665s6HRs	AGATGTACCTTCCACTTTCAC				
RgpB6H mutagene	esis primers				

 Table S3. Primers used for molecular biology work.

6HRgpBFA	CATCACCATCACCATCACTAATTCACACTGCAATTCTCTAATAAGG
DelRgpBFB	TAATTCACACTGCAATTCTCTAATAAGGGC
6HRgpBRA	GTGATGGTGATGGTGATGCTTCACTATAACCTTTTCTGTATACG
6HRgpBRB	CTTCACTATAACCTTTTCTGTATACGTCTTGC
Cloning primers	
PG266FrBXbalF	GCGTCTAGAGTATTCGGACTGCATCT
PG266FrBPstIR	ATTACTGCAGCGTCTGAATCGTGCAAA
PG266FrANdelF	TACCACATATGACTGTAGCAGCCATTACAC
PG266FrASmaR	ATTAACCCGGGCTCCTATGACTGCTTTGA
ermFAMXmalF	ATTACCCGGGATAGCTTCCGCTATTGC
ermFAMXbalR	GCTCTAGACGAAGCTGTCAGTAGTATACC
tetQXmalF	TATACCCGGGACAACGAATTATCTCCTTAACGT
tetQSallR	ATTTTGTCGACTTTTATTGCCAAGTTCTAATGCT

Figure S1. The structure-based alignment of the CTD of proteins secreted by *P. gingivalis* via T9SS. Conserved hydrophobic residues in β-strands are highlighted in grey. CTDs predicted from MS analysis of culture media are typed in blue with residues after which proteolysis occurred marked in a bold red font (Veith, P. D., *et al.* Protein substrates of a novel secretion system are numerous in the *Bacteroidetes* phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *J. Proteome Res.* **12**, 4449-4461 (2013)).

		sl	s2	s 3	s4	s5	<i>s6</i>	<i>s</i> 7
		.>>>>	>>>>>	>>>>>	>>>>.	· · · · · · · · · >>>>>>>>>>>>>>>>>>>>	>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	>>>>>>>>>
PG0506	VKVEG T SIADVANDKP	PYTVAVSGK	TITVESPA	-AGLTIFDMNG	RRVATAK-	NRMVFEA	-QNGVYAVRIAT	-EGKTYTE K VIVK
PG0182	ARFRLSYGCDENVDDS	SHVVSTNGR	EIIILNQDAL	DCTVTLFTIE G	KLLRRLK-	VLAGHREVMKVQT	G G AYIVHLQN	-AFTNDVH K VLVEY-
PG0183	FRLSYDEEWVESAEVS	SVLVGTAGK	RIVITNNSEH	ACQANVYTTD G	KLLIRLD-	VKPGSKSMTEPL	-VD G V Y VVSLQSI	PATSSNVR K VVVN
PG0193	RNRRQQELQDIQTRYQ	QSYQTMQE	DLQKRQQQ	-LFAPIQQKVA	DAIKKVG-1	DEENCAYIMEAGMML	-YT G ATAIDLTA	KVKA K LGIK
PG0232	ILDDSV E DIVAQTGIV	VIRPQNGTK	QILIEANAAI	-KAIVLYDING	RVVLKTT-	PNQLRSTVDLSI	LPE GIY TINIKT:	-EKSARTE K IHIG
PG0350	PYFPGITALISIEGES	SEYSVYAQD	GILYLSGMEQ	GLPVQVYTVG G	SMMYSSV-	ASGSAMEIQL	PRGAA Y VVRIG-	SHAI K TAMP
PG0410	ISISPNPAKAVVTIIY	YTDNPSCS	VIKIYG	-INGASADITG	LPKHLSE-	GYYSIQFNTSN	FDP G F Y LVTLNVI	DQKIIDTE K LRIK
PG0411	YTITSLDNIQSDTSLK	CIYPNPASY	VVRIEGLSRS	KSTIELYNAL G	ICILREE-	THSEKTEIDVSR	LND G V Y LIKVVG·	-GNKTTTE K VEIKRP
PG0495	KSNAVGEVDTKGFHVY	PIPTSKDL	TIEIPAEMVG	-KVASLIDMNG	QIVYRVT-	LNNIFQQIDISH	-LK G VFLLQIG-	DITE R VIVQ
PG0553	GIS <mark>N</mark> GVAQIENNNAVV	AYPSVVTD	RFSIKNAHMV	-HAAALYSLDG	KQVRSWN-	NLRNGVTFSVQG	LTA G T <mark>Y</mark> MLVMQT·	-ANGPVSQ K IVKQ
PG0611	TNCYPLSTKPVAGDDE	EVFVKQQGR	QIEIDSNSPI	-VQVVVYDLE G	KSVFRKR-	MTENAYTLSFRA	PML G FMTIMIET.	-QNSIINK K LNVTQL
PG0614	SCHTTDSQTVVPSSND	DINVYIQGT	TIGIKAEKLI	-KSVYIYDMAG	RMLFATS-	QTQGREFCIDL	KTK G HILVTVLFA	A-DNTQTS K NIIL
PG0616	IINAGQESLDKAEP <mark>T</mark> A	TEQIVATP	SVKAYVQ	-NGKIVVEY-S	KMEVFNA-	TGQLVKNES	LVP G V Y VVRITA-	-NGVMYFL K VLVP
PG0626	ILNYQSLQEVEQEGIR	RIYPNPTTD	RLMLDNVGDV	-SAIRIIDMYG	RPLVVVG-	NAGSSKLSIDVSD	LGR GNY VVDMQC·	-KGHRKIS R LITKE-
PG0654	MQG N ALTDVAVNESIK	CIYPRPATD	FLRIEGSQL-	-LRLSLFDMNG	KLIRATE-	LTGDLAIIGVAS	LPR G T <mark>Y</mark> IAEITAA	ANSKTIRA K VSLR
PG1030	KESFI T SFISPTVVQG	VDVYTLAG	KIRIESETPV	-SEVLLFDLAG	RMVLRQTI	DNKIYSDIDTNGLKR	SGIYVVSVRLS	SSGQVFSH K VQV
PG1326	NIVVANSANIYGADKF	PFALTVVGK	TIVASAFKG-	-EEITLYDIR G	RLIASGC-	DTLRYKA	-EN G F Y LIKIQV	-NGTVYTE K IQIQ
PG1374	SP T SNLAVDAPTVRIY	PNPVGRYA	LVEIPESLLG	-QEAALYDMNG	VKVYSFA-	VESLRQNIDLTH	LPD G T Y FFRLD-	NYTT K LIKQ
PG1424	TNTCTVTGAAKALRAW	FNAGRSEL	AVSVSLNIAG	TYRIKLYNTA G	EEVAAMTKI	ELVAGTSVFSMDVYS	QAP G T Y VLVVEG-	-NGIRETM K ILK
PG1427	GTAVEAIESSEEIRVF	PNPARDYV	EISAPCIPQE	-TSIILFDLS G	KIVMKNS-	LSAGHGRMDVSR	LPN G AYILKVD	GYTT K INIVH-
PG1548	SSSDIAGKDVSTIVLY	PNPAHDYV	HVAIPPTYAG	-STLRLFDIQG	RMQLSTK-	IESADMRLDVER	LPK G T Y IVVVE	DMVG K LFIR
PG1795	VHLKKG E GVEAVLTND	KANAYVQN	GVIYVAGANG	-RQVSLFDMNG	KVVYTGV-	SETIAA	PQKGM Y ILRVG-	AKSIK <mark>L</mark> AI
PG1798	KALTSMATPSTEAQVA	VYLNPSTD	RLVILANGI-	-THLSMYDLQG	KLIRDCA-	LSGDKVE <mark>M</mark> GVGS	LTK G T Y LLKVNT·	-DQGAFVR K VVIR
PG1837	ATLNI T SLADVTAQKP	PYTLTVVGK	TITVTCQ	-GEAMIYDMNG	RRLAAGR-	NTVVYTA	-QGGHYAVMVVV·	-DGKSYVE K LAVK
PG1844	ATLNI T SLADVTAQKP	PYTLTVVGK	TITVTCQ	-GEAMIYDMNG	RRLAAGR-	NTVVYTA	-QG G H Y AVMVVV·	-DGKSYVE K LAVK
PG2024	VDYIP <mark>D</mark> GVADVTAQKP	PYTLTVVGK	TITVTCQ	-GEAMIYDMNG	RRLAAGR-	NTVVYTA	-QGGYYAVMVVV·	-DGKSYVE K LAIK
PG2100	LDVGDVLQKEGSMKLY	PNPAKEYV	LINLPKEGG-	-HEAVVYDMQ G	RIVEKVS-	-FSGKEYKLNVQYLS	K G T Y MLKVVA·	-DTEYFVE K IIVE
PG2102	N GVEDIVMQEGSMKLY	PNPAQEYA	VISLPTAAN-	-CKAVVYDMQG	RVVAEAS-	FSGNEYRLNVQH	LAK G T <mark>Y</mark> ILKVVS·	-DTERFVE K LIVE
PG2172	KGGGTGLTNIGLGRIA	LIQSGNTC	TLQYNSNGKR	-LALEVYNLLG	VKVFTSQ-	-LPAGSGSYTLPVRL	-QRGVHIFRITE	GGKPAFVQ K YLIK
PG2198	VTVTN <mark>S</mark> SLSNVDGQAP	PYTLRVEGK	KIIAEAHG	-MEITLYDING	RTVAVAP-	NRLEYMA	-QT G F Y AVRFDV	-GNKHHVS K IQVR
PG2216	DNKPVI T SLAAPISHE	EIRIWATAG	RIFIAGAPAG	-TSVQVYDMQ G	HRIYNAA-	VLADHDIAVAS	G V Y VVRAG-	DSTA K VIVPPG
		-						
			iviotit A	Motif	В	Motif C	Motif D	Motif E

Figure S2. Superposition of the *ab initio* bead models and BUNCH atomistic models (cartoon representation of domains only) for the **A.** wild-type rIgSF-CTD and **B.** r664i6H variant. These models represent the best-fit average models to the data and show the two domains well separated in space with no significant interaction surface.



Β.



Figure S3. R_g (upper panel) and d_{max} (lower panel) distributions for the ensemble optimisation modelling (EOM) of wild-type rIgSF-CTD (black line) and r664i6H (red line) variant.



Figure S4. Comparison of RgpB phenotype expressed by wild-type *P. gingivalis* W83 and mutant strains with insertion (RgpB662iXa6H) or substitution (RgpB665sXa6H) of hexapeptide (IEGRAA) in the linker between IgSF and CTD domains. The whole culture (c) and spent growth medium (s) were subjected to Western blotting analysis using RgpB specific mAbs.

Figure S5. Guinier plots for the SAXS data for wild-type rlgSF-CTD ($qR_g < 1.18$, black symbols) and r664i6H ($qR_g < 1.15$, red symbols) variant. The solid lines are linear fits to the experimental data which shows the expected good fit for solutions of mono-disperse proteins.

