Supplemental material for:

Cpl-7, a Lysozyme Encoded by a Pneumococcal Bacteriophage with a Novel Cell Wall-Binding Motif

Noemí Bustamante[‡§], Nuria E. Campillo[¶], Ernesto García[§‖], Cristina Gallego[‡§], Benet Pera[‖], Gregory P. Diakun[**], José Luis Sáiz[‡§], Pedro García[§‖], J. Fernando Díaz[‖], and Margarita Menéndez[‡§1]

From the [‡]Instituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006, Madrid, Spain, the [§]CIBER de Enfermedades Respiratorias (CIBERES), the [¶]Instituto de Química Médica, Consejo Superior de Investigaciones Científicas, Juan de la Cierva 3, 28006, Madrid, Spain, the [‖]Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040, Madrid, Spain, and the [**]CCLRC Daresbury Laboratory, Warrington WA4 4AD, United Kingdom

Running head*: Cpl-7 pneumococcal phage endolysin*

[1] To whom correspondence should be addressed: Instituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006, Madrid, Spain. Tel.: 3491-561-9400; Fax: 3491-564-2431; E-mail: mmenendez@iqfr.csic.es.

Contents:

TABLE S1. Bacterial species encoding proteins containing CW_7 repeats

Bacteria or phage (host)[a]	Strain	Description (Evidence of phage origin)[b]	Protein containing CW_7 repeats	
			Accession No.[c]	Proposed function[d]
ACTINOBACTERIA		Gram-positive bacteria with a high G+C content in their DNA		
Actinomycetales		The pattern of 16S rRNA signatures consists of nucleotides at positions 688 : 699 (G–C), 701 (C), 823 : 877 (G–C) and 1060 : 1197 (U–A)		
Corynebacterium jeikeium	ATCC 43734	R, A, NSF. Human skin. Opportunistic, nosocomial pathogen (ϕ)	EEW16061 (1)	NAM-amidasa (2)
Mobiluncus curtisii	ATCC 43063	R, An, NSF. Genitourinary tract and rectum. Opportunist pathogen	EEI88064 (2)	NAM-amidase (2)
Mobiluncus mulieris	ATCC 35243	R, An, NSF. Genitourinary tract and rectum. Opportunistic pathogen (ϕ)	EEJ54470 (3)	NAM-amidase (2)
Propionibacterium acnes	SK137	R, An, NSF. Can tolerate oxygen. Oral cavity, large intestine, conjunctiva, skin. Causes acne vulgaris as well as postoperative and device-related infections	ADD99701 (4)	NAM-amidase (2)
Bifidobacteriales		The pattern of 16S rRNA signatures consists of nucleotides at positions 688 : 699 (A–U), 701 (U), 823 : 877 (A–U) and 1060 : 1197 (C–G)		
Bifidobacterium adolescentis	L2-32	R, An, NSF. Gastrointestinal tract of adults	EDN82563 (5)	Lysozyme
		(ϕ)	EDN82835 (6)	Lysozyme
Bifidobacterium dentium	ATCC 27678	R, An, NSF. Oral cavity. Causes human dental caries (ϕ)	EDT45852 (7)	Lysozyme
	Bd1		ADB10077	
Bifidobacterium longum	DJO10A	R, An, NSF. Gastrointestinal tract of adults (ϕ)	ACD98581 (8)	Lysozyme
Bifidobacterium longum subsp. *infantis*	ATCC 15697	R, An, NSF. Gastrointestinal tract of infants	ACJ52624 (9)	NK[1]
			ACJ52650 (10)	Lysozyme
		(ϕ)	ACJ52906 (11)	Lysozyme
Bifidobacterium longum subsp. *longum*	F8	R, An, NSF. Gastrointestinal tract of adults (ϕ)	CBK71086 (12)	Lysozyme
Coriobacteriales		The pattern of 16S rRNA signatures consists of nucleotides at positions 242 : 284 (C–G), 291 : 309 (C–G), 316 : 337 (U–G), 819 (A), 952 : 1229 (U–A) and 1115 : 1185 (C–G)		
Collinsella intestinalis	DSM 13280	R, An, NSF. Human intestine (ϕ)	EEP44864 (13)	Lysozyme
Collinsella stercoris	DSM 13279	R, An, NSF. Human intestine (ϕ)	EEA89601 (14)	Lysozyme
		(ϕ)	EEA90700 (15)	Lysozyme
BACTEROIDETES		The *Bacteroidetes* species are presently distinguished from other bacteria solely on the basis of their branching patterns in the 16S rRNA trees. Gram-negative, anaerobic rods		
Bacteroidales				
Bacteroides pectinophilus	ATCC 43243	Human intestine (ϕ)	EEC56840 (16)	Glucosaminidase
Bacteroides stercoris	ATCC 43183	Human intestine	EDS16215 (17)	NK
Porphyromonas uenonis	60-3	Human intestine	EEK16248 (18)	NK
Prevotella copri	DSM 18205	Human intestine	EFB34745 (19)	NK
CHLOROFLEXI		Mostly Gram-negative, filamentous bacteria with unusual cell envelopes		
‘*Dehalococcoidetes*’		Chloroethene-respiring bacteria		
‘*Dehalococcoides ethenogenes*’	195	Contaminated groundwater sites (ϕ)	AAW39696 (20)	Lysozyme + CHAP
FIRMICUTES		Gram-positive bacteria with a low G+C content in their DNA		
Clostridiales		This order is phenotypically, chemotaxonomically, physiologically and ecologically diverse. Contains both Gram-positive and Gran-negative organisms		
Anaerococcus hydrogenalis	DSM 7454	C, An, NSF. Vaginal discharges and ovarian abscesses (ϕ)	EEB35354 (21)	NAM-amidase (2)
Anaerococcus lactolyticus	ATCC 51172	C, An, NSF. Vaginal discharges and ovarian abscesses	EEI85350 (22)	Lysozyme
		(ϕ)	EEI86118 (23)	Gucosaminidase
Anaerococcus prevotii	DSM 20548	C, An, NSF. Vaginal discharges and ovarian, peritoneal and sacral abscesses	ACV28608 (24)	Lysozyme
Anaerococcus tetradius	ATCC 35098	C, An, NSF. Vaginal discharges and ovarian abscesses	EEI83902 (25)	Lysozyme

Finegoldia magna	ATCC 53516	C, An, NSF. Gastrointestinal and female genito-urinary tract. Causes endocarditis and other infections (ϕ)	EEJ67759 (26)	NAM-amidase (2)
Parvimonas micra	ATCC 33270	C, An, NSF. Oral cavity and gastrointestinal and female genito-urinary tract. Endodontic abscesses	EDP23940 (27)	NAM-amidase (5)
Anaerostipes caccae	DSM 14662	Gram-variable, R, An, NSF. Human intestine (ϕ)	EDR98670 (28)	NAM-amidase (2)
Catonella morbi	ATCC 51271	Gram-negative, R, An, NSF. Oral cavity. Endodontic infections	EEP22199 (29)	CHAP
		(ϕ)	EEP22813 (30)	CHAP
		(ϕ)	EEP22668 (31)	Glucosaminidase + CHAP
Clostridium sp.	M62/1	R, An, SF. Human intestine (ϕ)	EFE12573 (32)	Lysozyme
Clostridium asparagiforme	DSM 15981	R, An, SF. Human intestine (ϕ)	EEG51680 (33)	NK
Clostridium hathewayi	DSM 13479	R, An, SF. Human intestine	EFC95213 (34)	CHAP
Clostridium leptum	DSM 753	R, An, SF. Human intestine	EDO61912 (35)	NAM-amidase (2) + sporulation/cell division
Coprococcus catus	GD/7	C, An, NSF. Human intestine (ϕ)	CBK80688 (36)	Lysozyme
Coprococcus comes	ATCC 27758	C, An, NSF. Human intestine	EEG89528 (37)	NK
		(ϕ)	EEG89729 (38)	Lysozyme
Coprococcus eutactus	ATCC 27759	C, An, NSF. Human intestine	EDP25145 (39)	Lysozyme
Ethanoligenens harbinense	YUAN-3	R, An, NSF. Wastewater (ϕ)	EFD38839 (40)	Lysozyme
Eubacterium rectale	ATCC 33656	R, An, NSF. Human intestine (ϕ)	ACR75820 (41)	Lysozyme
Ruminococcus gnavus	ATCC 29149	R, An, NSF. Human intestine (ϕ)	EDN76657 (42)	Glucosaminidase + CHAP
		(ϕ)	EDN79250 (43)	Lysozyme
		(ϕ)	EDN79487 (44)	NK
Ruminococcus obeum	ATCC 29174	C, An, NSF. Human intestine (ϕ)	EDM87532 (45)	NK
Ruminococcus torques	L2-14	C, An, NSF. Human intestine (ϕ)	CBL24890 (46)	Glucosaminidase
Clostridiales bacterium sp.	SS3/4		CBL42507 (47)	NAM-amidase (2) + Glucosaminidase + sporulation/cell division
Erysipelotrichales		Nonmotile, slender rods with a Gram-positive cell-wall structure. Aerobic to facultatively anaerobic		
Bulleidia extructa	W1219	R, An, NSF. Oral cavity. Dental abscesses (ϕ)	EFC05584 (48)	Lysozyme
Erysipelothricaceae bacterium	5_2_54FAA	Human microbioma (ϕ)	EFE46538 (49)	NAM-amidase (2)
Lactobacillales		Gram-positive rods and cocci. Endospores are not formed. Usually facultatively anaerobic and catalase-negative		
Granulicatella elegans	ATCC 700633	R, An, NSF. Oral cavity. Endocarditis (ϕ)	EEW92387 (50)	NAM-amidase (2)
Lactobacillus antri	DSM 16041	R, M, NSF. Human stomach mucosa (ϕ)	EEW53895 (51)	Lysozyme
Lactobacillus gasseri	ATCC 33323 202-4	R, An, NSF. Human intestine (ϕ)	ABJ60844 (52) EEQ25513	Lysozyme
Lactobacillus reuteri	SD2112 CF48-3A	R, An, NSF. Human intestine (ϕ)	EEJ91939 (53) EEI65431	NK
Lactococcus lactis subsp. *lactis*	IL1403	C, A, NSF. Dairy products (ϕ)	AAK05760 (54)	Transglycosylase
	KF147	C, A, NSF. Plant material (ϕ)	ADA65407 (55)	Transglycosylase
Lactococcus lactis subsp. *cremoris*	SK11	C, A, NSF. Dairy products (ϕ)	ABJ73318 (56)	Transglycosylase
	MG1363	C, A, NSF. Dairy products (ϕ)	CAL97364 (57)	Transglycosylase
Streptococcus dysgalactiae subsp. *equisimilis*	GGS_124	C, FA, NSF. Respiratory tract and female genito-urinary tract. Pharyngitis, glomerulonehritis, and septicemia. (ϕ)	BAH82103 (58)	NAM-amidase (5) + glucosaminidase
Phages				
Cp-7 (*S. pneumoniae*)		Lytic phage	AAA72844 (59ϕ)	Lysozyme

λSa2 prophage(*S. agalactiae* 2603 V/R)	Prophage	AAN00700 (60ϕ)	NAM-amidase (5)[e] + glucosaminidase
ϕ10750.3 prophage (*S. pyogenes* MGAS10750)	Prophage	ABF38229 (61ϕ)	NAM-amidase (5) + glucosaminidase
ϕSpeA prophage (*S. pyogenes* MGAS8232))	Prophage	AAL97140 (62ϕ)	NAM-amidase (5) + glucosaminidase
ϕ315.3/SPsP4/M1T1.Y prophage(*S. pyogenes* MGAS315/SSI-1/M1T1)	Prophage	AAM79703 (63ϕ) BAC63864 AAX92493	NAM-amidase (5) + glucosaminidase
ϕMan.2 prophage (*S. pyogenes* Manfredo)	Prophage	CAM30020 (64ϕ)	NAM-amidase (5) + glucosaminidase
SMP (*S. suis*)	Lytic phage	ABK91888 (65ϕ)	NAM-amidase (5) + glucosaminidase
Bbr-1 prophage (*Bifidobacterium breve* UCC 2003)	Prophage	AAY16487 (66ϕ)	NK
KSY1 (*L. lactis*)	Lytic phage	ABG21617 (67ϕ)	Transglycosylase

[a] Phyla (in uppercase lettering), order (italicized and underlined), and species names were assigned following current rules (1) as well as the Euzéby's List of Prokaryotic Names with Standing in Nomenclature (http://www.bacterio.cict.fr/). Quotation marks indicate taxonomic categories not covered by the Rules of the Bacteriological Code.

[b] Relevant characteristics are abbreviated as follows: A, aerobic; An, anaerobic, C, coccus; FA, facultative anaerobic; M, microaerophylic; NSF, non-spore-forming; R, rod; SF, spore-forming. ϕ indicates the presence of a holin-like coding gene (or other phage-related genes) located immediately upstream of that encoding a protein containing CW_7 repeats. This finding was considered as an indication of phage origin.

[c] Accession numbers correspond to those included in the Entrez Protein Database of the National Center for Biotechnological Information (http://www.ncbi.nlm.nih.gov/). Numbers in parentheses correspond to those of Fig. 1.

[d] Proposed functions according with the Pfam protein family database (version 24.0). NK, not known.

[e] γ-D-glutaminyl-L-lysine endopeptidase activity has been demonstrated for the *Amidase_5* motif of the λSa2 prophage (2).

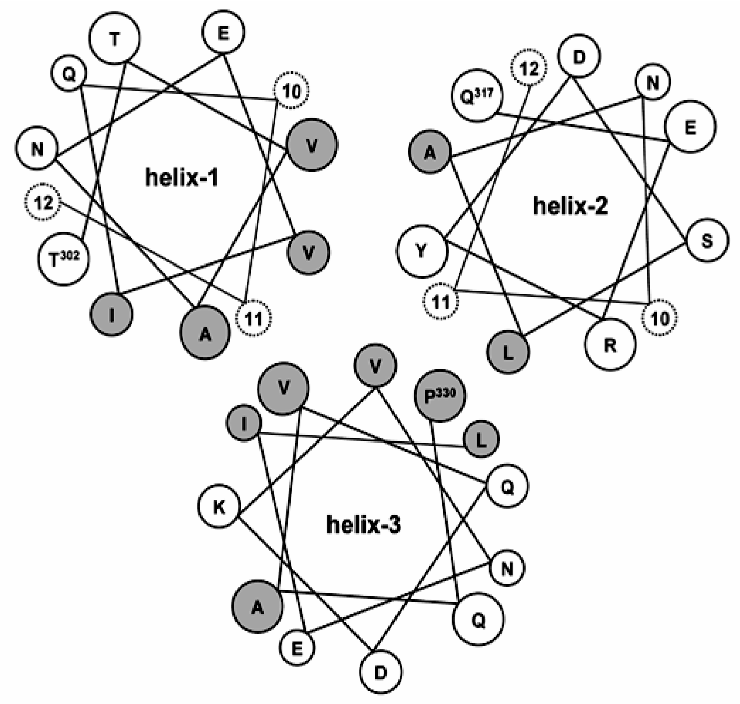
REFERENCES

1. Euzéby, J. (2010) *Int. J. Syst. Evol. Microbiol.* **60,** 469–4721
2 Pritchard, D. G., Dong, S., Kirk, M. C., Cartee, R. T., and Baker, J. R. (2007) *Appl. Environ. Microbiol.* **73**, 7150-7154

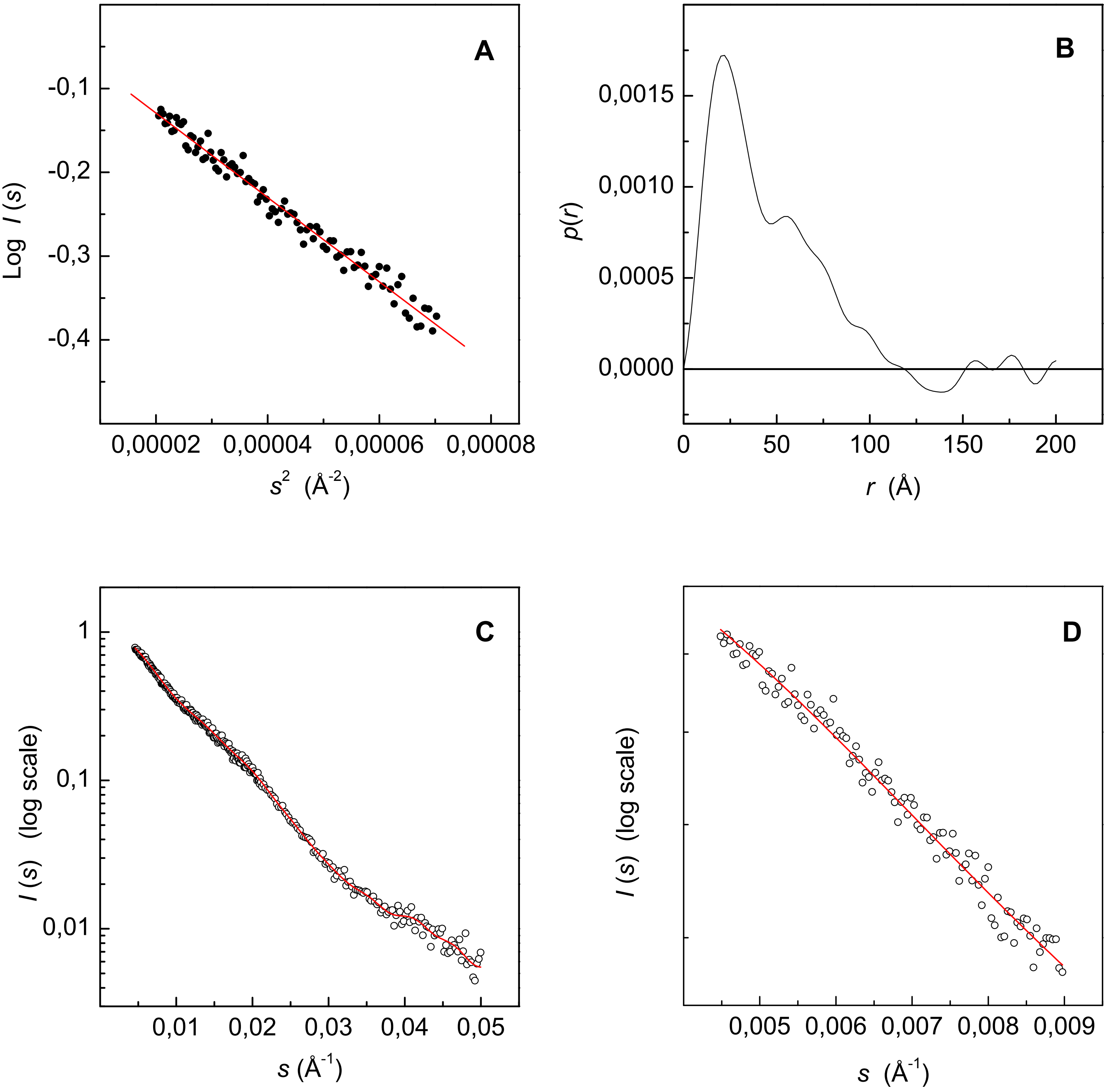
Legend for Supplemental Figures

Figure S1 – Helical wheel diagram of the three α-helices predicted for each CW_7 repeat. The first residue in the wheels is numbered according to the third Cpl-7 repeat (Thr302, Gln317 and Pro330); hydrophobic residues at the internal faces are filled in gray. Dotted circles are unoccupied positions in the wheel.

Figure S2 – *A*, Guinier plot of SAXS data. *B*, extended plot of the *P*(*r*) function. *C*, fit of the Fourier transform of the *P*(*r*) function to SAXS data. *D*, fit of theoretical scattering profile of the *ab initio* model of Cpl-7 generated by DALAI-GA to the SAXS data at low values of the scattering vector (*s*).



Bustamante et al. Supplementary Fig. S1



Bustamante *et al.* Supplementary Fig.S2