PLOS PATHOGENS

RESEARCH ARTICLE

Designed Ankyrin Repeat Proteins provide insights into the structure and function of CagI and are potent inhibitors of CagA translocation by the *Helicobacter pylori* type IV secretion system

Marine Blanc[1], Clara Lettl[2], Jérémy Guérin[1], Anaïs Vieille[1], Sven Furler[3], Sylvie Briand-Schumacher[3], Birgit Dreier[3], Célia Bergé[1], Andreas Plückthun[3], Sandrine Vadon-Le Goff[4], Rémi Fronzes[5], Patricia Rousselle[4], Wolfgang Fischer[2]*, Laurent Terradot[1]*

1 UMR 5086 Molecular Microbiology and Structural Biochemistry CNRS-Université de Lyon, Institut de Biologie et Chimie des Protéines, Lyon, France, **2** Max von Pettenkofer Institute of Hygiene and Medical Microbiology, Faculty of Medicine, LMU Munich, Munich, Germany, **3** Department of Biochemistry, University of Zurich, Zurich, Switzerland, **4** University of Lyon, CNRS UMR5305, Tissue Biology and Therapeutic Engineering Laboratory (LBTI), Lyon, France, **5** European Institute of Chemistry and Biology, CNRS UMR 5234 Microbiologie Fondamentale et Pathogénicité, Univ. Bordeaux, Pessac, France

* Fischer@mvp.lmu.de (WF); laurent.terradot@ibcp.fr (LT)

Citation: Blanc M, Lettl C, Guérin J, Vieille A, Furler S, Briand-Schumacher S, et al. (2023) Designed Ankyrin Repeat Proteins provide insights into the structure and function of CagI and are potent inhibitors of CagA translocation by the *Helicobacter pylori* type IV secretion system. PLoS Pathog 19(5): e1011368. https://doi.org/10.1371/journal.ppat.1011368

Editor: Tomoko Kubori, Gifu University, JAPAN

Received: November 9, 2022

Accepted: April 18, 2023

Published: May 8, 2023

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The project was funded by ANR grants ANR-Subsist (ANR- 19-CE11-0012). JG salary was provided by ANR- 19-CE11-0012. The study was also funded by ANR Sintesys grant ANR-13-ISV3-0006, which provided salary to CB. This work also benefited from support from the Ligue Régionale contre le Cancer, grant number 2017 and Pack

Abstract

The bacterial human pathogen *Helicobacter pylori* produces a type IV secretion system (*cag*T4SS) to inject the oncoprotein CagA into gastric cells. The *cag*T4SS external pilus mediates attachment of the apparatus to the target cell and the delivery of CagA. While the composition of the pilus is unclear, CagI is present at the surface of the bacterium and required for pilus formation. Here, we have investigated the properties of CagI by an integrative structural biology approach. Using Alpha Fold 2 and Small Angle X-ray scattering, it was found that CagI forms elongated dimers mediated by rod-shape N-terminal domains ($CagI^N$) prolonged by globular C-terminal domains ($CagI^C$). Three Designed Ankyrin Repeat Proteins (DARPins) K2, K5 and K8 selected against CagI interacted with $CagI^C$ with subnanomolar affinities. The crystal structures of the CagI:K2 and CagI:K5 complexes were solved and identified the interfaces between the molecules, thereby providing a structural explanation for the difference in affinity between the two binders. Purified CagI and $CagI^C$ were found to interact with adenocarcinoma gastric (AGS) cells, induced cell spreading and the interaction was inhibited by K2. The same DARPin inhibited CagA translocation by up to 65% in AGS cells while inhibition levels were 40% and 30% with K8 and K5, respectively. Our study suggests that $CagI^C$ plays a key role in *cag*T4SS-mediated CagA translocation and that DARPins targeting CagI represent potent inhibitors of the *cag*T4SS, a crucial risk factor for gastric cancer development.

Ambition Rhône-Auvergne Région 2020 grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Helicobacter pylori is a bacterial pathogen that colonises the human stomach in half of the world's population. The most virulent strains use the *cag*- type IV secretion system (*cag*T4SS), a molecular nanomachine capable of injecting the oncoprotein CagA into gastric cells. How CagA is delivered is unknown, but the *cag*T4SS apparatus contains several proteins that can interact with host cell receptors, mediating CagA translocation from the cytoplasm of the bacteria to the inner leaflet of the host cell membrane. In this study we have investigated the structural and functional properties of CagI, a protein long-known to be associated with the *cag*T4SS apparatus but with yet unknown function. We found that CagI displays a unique dimeric structure and that its C-terminal domain is involved in interaction with the host cell. Designed Ankyrin Repeat Proteins were selected against CagI and found to interact with its C-terminal moiety with high affinity. DARPin binding was able to prevent CagI interaction with the host cell and inhibited CagA translocation by *H. pylori*. Our study reveals the role of CagI in *cag*T4SS interaction with gastric cells and provides a first example of a small protein binder inhibiting the *cag*T4SS activity.

Introduction

Helicobacter pylori is a Gram-negative bacterium that colonizes the human stomach in half of the world's population and it is a major risk factor for the development of gastric diseases, including ulcers and gastric cancers [1]. Strains carrying the *cag*-pathogenicity island (*cag*PAI) are more frequently associated with severe diseases [2]. This *cag*PAI is a 40 kbp DNA region that encodes for a type IV secretion system (*cag*T4SS) and for the CagA oncoprotein. Upon contact with gastric cells, the *cag*T4SS delivers CagA into epithelial gastric cells. Once injected, CagA attaches to the inner leaflet of the membrane of the cell where it can be phosphorylated by host cell kinases and interacts with a plethora of cell signalling proteins, hereby promoting tumor development [3]. Although CagA is the only protein effector, several molecules have been reported to be translocated by the *cag*T4SS machinery, including DNA, peptidoglycan, or ADP-heptose that have important pro-inflammatory effects [4].

T4SSs are versatile multi-protein bacterial devices used to transport macromolecules across membranes in various biological processes such as natural transformation, conjugation or delivery of protein effectors into target cells [5,6]. Based on the prototypical T4SS VirB/D from *Agrobacterium tumefaciens*, T4SSs consist of 12 proteins, VirB1-B11 and VirD4, that form a molecular nanomachine. T4SSs generally comprise a core machinery, made of heteromultimers of VirB3-10 that form a stable complex spanning both bacterial membranes [7]. This core machine is used to produce the external pilus and to translocate substrates [5]. Pilus biogenesis requires the recruitment of VirB11 to VirB4 that promotes the assembly of VirB2 subunits and phospholipids capped by the minor pilin VirB5 [8]. The delivery of substrates relies on the VirD4 ATPase that serves as a coupling protein with the core machinery by interacting with VirB10 [9]. The *cag*PAI encodes for around 27 proteins, some of which are clear homologues of VirB/D proteins but others have no homologues outside *H. pylori* genomes [10]. As a consequence, the core structure of the machinery comprises several additional proteins and is unusually large [4,11]. Early studies using scanning electron microscopy found that the *cag*T4SS pilus appeared as a flexible sheathed structure protruding outside the cells [12,13]. Similar tubular structures containing lipopolysaccharides that were located nearby *cag*T4SS complexes were more recently imaged by cryo-electron tomography [14]. Nevertheless, the

*cag*T4SS pilus composition is unclear and CagC, the homologue of the major pilin VirB2, was found to be dispensable for pilus formation [15].

The production of a functional pilus requires CagL, CagI and CagH, three proteins specific to the *cag*T4SS that are located on the same operon [16]. CagI, CagL and CagH might interact together, and each of them is required for CagA translocation [16,17]. Much attention and studies have focused on CagL (reviewed in [18]), since the protein mediates cell attachment *in vitro* and possesses an arginine-glycine-aspartate (RGD) motif involved in integrin binding [19–22]. CagL structures [19,23–25] revealed a six-helix bundle with a cysteine-clasped loop important for its function [26]. Despite having no clear structural homology, CagL is considered a functional homologue of VirB5 since it shares biophysical properties, it is located at the tip of the pilus and interacts with host cell receptors [19,27]. However, CagL is not the only Cag protein able to bind to integrins [18] and thus multiple interactions might take place at the host-bacterium interface. CagI interacts with integrin $\alpha_5\beta_1$ *in vitro* and *in vivo* [28,29] and is detected at the surface of *H. pylori* cells [16,30,31]. CagI might be also directly associated with the *cag*T4SS core-complex assembly, since deletion of the *virB8* homologue *cagV*, *cagX* and *cagY* resulted in CagI instability [31,32]. The C-terminus of CagI shows sequence conservation with CagL including the two cysteines positioned approximately 100 residues upstream of the C-terminal residues [19] and a C-terminal hexapeptide motif essential for CagA secretion [16]. CagI interacts with CagL *in vitro* [28] and influences its stability *in vivo* [31].

Here, we have used an integrated structural biology approach to gain insights into the structure of CagI. Our data reveal that CagI is a modular protein consisting of an N- and C-terminal domain. The N-terminal domain mediates the protein dimerization and the C-terminal domain is able to mediate cell adhesion and spreading. We also selected Designed Ankyrin Repeat Proteins (DARPins, [33]) against the CagI protein to probe its function. DARPins are small (14–18 kDa), highly stable, α-helical scaffolds that can bind with high affinity to their targets and have thus applications in various fields, including crystallography, diagnostics and therapeutics [33]. Each repeat consists of 33 amino acids, of which 7 are randomized. Two or three internal repeats are stacked and are flanked by N- and C-terminal capping repeats, to result in N2C or N3C structures. Three DARPins targeting the C-terminal domain of CagI were found to inhibit CagI-mediated cell adhesion and CagA translocation in human cells by up to 65%. These results point towards a key role for CagI in the injection of the main *cag*T4SS effector, possibly by facilitating pilus adhesion to the host cell receptors, thereby identifying a potential way to inhibit cagT4SS and help control *H. pylori*-mediated oncogenesis.

Results

CagI forms elongated dimers assembled via the N-terminal region

To investigate the structure of CagI, the protein was purified without its predicted signal peptide (corresponding to residues 1–20) and its molecular mass determined by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). CagI eluted as a single peak with a mass of 79 kDa, consistent with a dimer (Fig 1A). We then used AlphaFold 2 (AF) [34] to predict the CagI dimer structure and generated three different models (S1 Fig). The monomers of CagI in the three models showed an all α-helical structure consisting of two defined domains (Fig 1B). The N-terminal region ($CagI^N$) comprising residues 21 to 190 consists of two extended helices (α1 and α2) forming a helical hairpin, followed by a short helix α3 (Fig 1B). α3 connects $CagI^N$ to the protein C-terminal domain ($CagI^C$) encompassing residues 191 to 381. $CagI^C$ is a globular domain made of a four-helix bundle (α4-α7), reminiscent with CagL structure (see below). $CagI^C$ models showed some variations in the orientation of α7 but also in the conformation of α6 that was split in two helices with a kink between residues 283

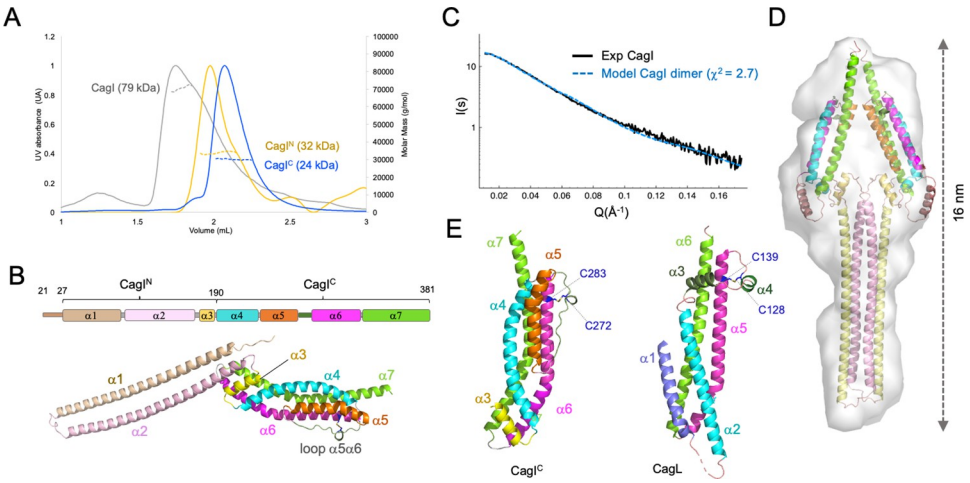


Fig 1. Integrative structural biology study of CagI. A) Size exclusion chromatograms (A_{280}) of CagI, $CagI^N$ and $CagI^C$. MALS weight-averaged molar masses are indicated as dotted lines. B) Schematic representation of CagI predicted secondary structures (top) and cartoon representation of Alpha Fold (AF) model of the CagI monomer with helices coloured as in the schematic view. C) Comparison of CagI dimer theoretical SAXS curve with experimental curve. D) Cartoon depiction of the AF model of CagI dimer coloured as in A) fitted in the SAXS envelope obtained with DAMMIF. E) Comparison of $CagI^C$ and CagL (PDB ID: 4YVM) depicted as cartoon. CagI is coloured as in A). CagL secondary structure elements equivalent of those of CagI are coloured accordingly. Cysteine residues involved in disulfide bridges are coloured in dark blue and displayed as ball-and-stick.

https://doi.org/10.1371/journal.ppat.1011368.g001

and 286 in one model. CagI structures were predicted with low predicted per-residue confidence score (pLDDT) of 30 to 50, except for residues 205 to 290 that were nearly identical in all models and encompassed helix α3 to half of α6 (S1 Fig).

AF predicted that the helical hairpin of the N-terminal region was involved in coiled-coil association in two out of three CagI dimer models (S1 Fig). To evaluate the contribution of the N- and C-terminal domains to CagI oligomerisation, we produced them separately and used SEC-MALS to measure their molecular mass (Fig 1A). $CagI^N$ (17 kDa) was found to have a mass of 32 kDa, thus consistent with a dimer in solution. The mass of $CagI^C$ (21.5 kDa) measured by MALS was 24 kDa, demonstrating that the isolated domain was monomeric in solution. To gain insight into the overall architecture of the protein in solution, we turned to size exclusion chromatography coupled to small angle X-ray scattering (SEC-SAXS) and collected data on CagI, $CagI^N$ and $CagI^C$. Comparison of the experimental SAXS curve of CagI with the theoretical ones obtained with AF models showed that model 3 and 2 display χ^2 values of 2.7 and 3.8, respectively, while model 1 showed the highest value 10.2 (Figs 1C and S2). Fitting the dimer model 3 in the *ab initio* DAMMIF envelope confirmed the general shape of the CagI dimer (Fig 1D). Similar experiments performed with the individual domains $CagI^N$ and $CagI^C$ confirmed that model 3 fitted better (S2 Fig). Molecular weight calculation using SAXSMoW2 [35] confirmed that $CagI^N$ forms a dimer and $CagI^C$ is monomeric, in agreement with SEC--MALS data (S1 Table). Thus we concluded that, CagI is a dimer in solution mediated by a head-to-head coiled-coil of N-terminal extended α1-α2 hairpins followed by individual $CagI^C$ domains.

CagI was proposed to have homology to the other pilus-associated protein CagL based on sequences and motif similarities [25]. Structural superimposition of CagI model with the CagL structure (PDB code 4YVM, [23]) shows that the similarity is limited to CagI helix α6 and the beginning of loop α5-α6 with helix α5 and loop α4-α5 in CagL, respectively (Fig 1E). This includes a conserved disulfide bridge between C272 and C283 predicted in CagI that tethers the α5-α6 loop to α6, corresponding to C128 and C139 in CagL (Figs 1E and S3). The

remaining parts of CagL structure and $CagI^C$ are rather different with, notably the two short helices α3 and α4 of CagL absent in the CagI structure (Figs 1B, 1E and S3) and an additional helix α5 in CagI.

DARPins against *Helicobacter pylori* CagI bind with high affinity to the C-terminal domain of the protein

To generate DARPin binders against CagI, avi-tagged CagI protein ($CagI_{avi}$) was immobilized alternatingly on streptavidin and neutravidin. Ribosome display selections were performed over four round using a semi-automated 96-well plate format (see Materials and Methods for the detailed procedure). After the fourth round of ribosome display selection, 380 single DARPin clones were expressed with N-terminal $MRGS(H)_8$ and C-terminal FLAG tag and screened for binding to CagI with a C-terminal strep-tag ($CagI_{strep}$) using ELISA on streptactin-coated plates. From the initial hits of the 380 analyzed DARPin clones, 32 were randomly chosen for further analysis and their sequence determined. Of these, 15 were identified as unique clones and they were expressed in a 96-well format and IMAC purified. The purified DARPins were used in a hit validation to bind immobilized $CagI_{strep}$ by ELISA using FLAG detection.

Next, we investigated if co-expression of the DARPins with CagI in *E. coli* cells could lead to co-purification. The vector expressing $CagI_{strep}$ was co-transformed in *E. coli* cells with each vector expressing a DARPin (K1 to K15) fused to an N-terminal His_8-tag. To monitor complex formation, we purified cell extracts on Ni-NTA beads and determined by SDS-PAGE if $CagI_{strep}$ co-purified with the DARPin. We observed that $CagI_{strep}$ co-eluted with DARPin K2, K5, K8, K9, K10, K11, K12 and K15 (S4 Fig). The presence of $CagI_{strep}$ was assessed by Western bloting using antibodies against the strep-tag (S4 Fig) in all elution fractions that were positive in SDS-PAGE. A band corresponding to $CagI_{strep}$ was also detected in the elution fraction of $CagI_{strep}$:K1 although no band was visible in the SDS-PAGE. The amino-acid sequences of the confirmed DARPins are shown in S5 Fig.

We then purified His_8-tagged DARPins K2, K5, K8-K12 and K15 and confirmed that they could also bind $CagI_{strep}$ in pull down assays on Ni-NTA beads (Fig 2A). The same assays performed with individual CagI domains showed that all DARPins interacted with $CagI^C$ but not with the $CagI^N$ (Fig 2A). To determine the affinity of the binders for their target, surface plasmon resonance (SPR) experiments were performed by immobilizing CagI or $CagI^C$ on the chip and injecting increasing concentrations of DARPins. DARPins showed different binding modes and K_D ranges for CagI and $CagI^C$ (Figs 2B, 2C and S6 and Table 1). For measurements on full-length CagI, best fits were obtained with the heterogenous ligand interaction model (S6A Fig), while interactions of DARPins with $CagI^C$ fit well the 1:1 binding model (Figs 2C and 6B). This suggests that DARPins interact with the C-terminal domain of CagI in a 1:1 manner, and that CagI dimerization reduces the accessibility of the two C-terminal domains. Because K_D values for CagI-DARPins measurements were similar between 1:1 binding and heterogenous ligand models, we considered K_D values obtained using the 1:1 binding model to compare the affinities between CagI and $CagI^C$ (Table 1). The measured K_D's indicated that DARPins had high affinity for CagI in the range of 1–10 nM except for K15 whose K_D was 73.7 nM. DARPins K2 and K8 had particularly low k_{off}, resulting in a strong affinity for CagI. For all the DARPins, affinities for $CagI^C$ were ten times higher, in the range of 0.2–1 nM, while for K2, K8 and K11 K_D's as low as 0.03 nM, 0.06 nM and 0.09 nM, respectively, were obtained. K15 again showed a weaker K_D of 4.89 nM (S6 Fig and Table 1). Interestingly, sequences of DARPin K2 and K5 differed only by a single amino-acid at position 125, being a leucine for K5 and a phenylalanine for K2 suggesting that the binding modes of the two DARPins were similar but with a 10-fold increase in affinity for K2 compared to K5.

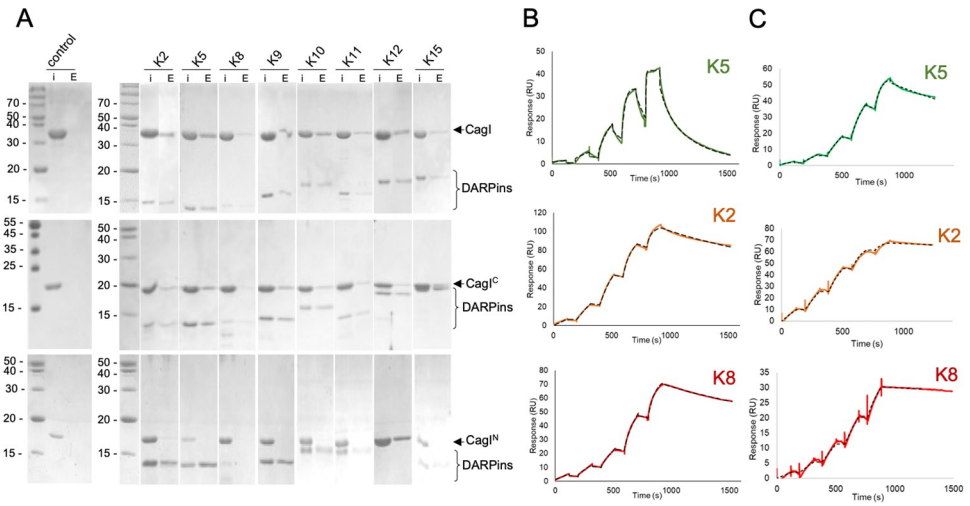


Fig 2. DARPin interactions with CagI. A) Pull down assays of purified untagged CagI (top panels), CagI^{C} (middle panels) or CagI^{N} (bottom panels) with NTA bead-immobilized His$_8$-tagged DARPins. "I" denotes input protein and "E" denotes elution. In control experiments proteins were mixed with the resin in the absence of DARPin and were not detected in the elution fraction. B) Representative SPR experiments using single-cycle injection mode on CM5 chips coated with CagI or with C) CagI^{C}. DARPins K5 (green curves), K2 (orange curves) or K8 (red curves) were injected on the chips at increasing concentrations as follows. For full-length CagI experiments, concentrations were 0.5, 2.5, 12.5, 62.5 and 312.5 nM for K5 and K8. For K2, concentrations were 1, 3, 9, 27 and 81 nM. For CagI^{C} experiments, K2 and K8 were injected at 0.05, 0.15, 0.45, 1.35 and 4 nM. For K5 concentrations used were 0.11, 0.33, 1, 3 and 9 nM. Fit curves obtained with binding model 1:1 are shown as dashed lines.

https://doi.org/10.1371/journal.ppat.1011368.g002

Structures of CagI:K5 and CagI:K2 complexes

To better understand the molecular basis of DARPin interaction with CagI, we solved the crystal structures of the CagI:K2 and CagI:K5 complexes. The two crystals had very similar cell parameters (see Table 2 for data collection and refinement statistics). The two complex structures revealed that each asymmetric unit contained one DARPin molecule and a fragment of CagI that had undergone proteolysis during crystallisation (Fig 3A). Density was clear for residues 204–307 of CagI in the two structures and these were nearly identical, with a rmsd of 0.2 Å^2. The structure of the fragment of CagI consists of a three-helix bundle corresponding to α4, α5 and α6 and the extended α5-α6 loop containing a 3_{10} helix in the AF model (Fig 3A and 3B). Interestingly, the fragment of the CagI crystallised corresponds to the region of the AF model that showed the highest prediction scores. Comparison of the structure of CagI$^{205\text{-}304}$

Table 1. Dissociation constant (K_D) expressed in nM obtained in Surface Plasmon Resonance experiments with immobilized CagI or CagI^{C} and the DARPins as analytes. Values were obtained using the 1:1 binding model (mean of two separate experiments except for values labeled with a *, for which a single multi-injection experiment was performed).

	CagI				CagI^{C}			
	K_D (nM)	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	χ^2	K_D (nM)	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	χ^2
K2	1.59 ± 0.91	$1{\cdot}10^{5}$	$4{\cdot}10^{-4}$	3–9	0.03 ± 0.01	$2{\cdot}10^{6}$	$2{\cdot}10^{-4}$	0.05–1
K8	1.15 ± 0.21	$1{\cdot}10^{5}$	$2{\cdot}10^{-4}$	1–10	0.06 ± 0.001	$1{\cdot}10^{6}$	$8{\cdot}10^{-5}$	0.05–0.3
K11	1.15 ± 0.02	$8{\cdot}10^{4}$	$1{\cdot}10^{-3}$	25–75	0.09 ± 0.01	$4{\cdot}10^{6}$	$4{\cdot}10^{-4}$	3.4–4.6
K10	3.83 ± 0.37	$8{\cdot}10^{4}$	$5{\cdot}10^{-3}$	9–62	0.41 ± 0.10	$5{\cdot}10^{6}$	$6{\cdot}10^{-4}$	2.5–6.3
K9	2.80*	$2{\cdot}10^{6}$	$4{\cdot}10^{-3}$	44	0.22 ± 0.01	$2{\cdot}10^{6}$	$6{\cdot}10^{-4}$	0.3–1.2
K5	8.05*	$2{\cdot}10^{5}$	$1{\cdot}10^{-3}$	0.8	0.23 ± 0.01	$2{\cdot}10^{6}$	$5{\cdot}10^{-4}$	0.2–1.7
K12	8.61*	$2{\cdot}10^{5}$	$1{\cdot}10^{-3}$	43	0.81 ± 0.34	$1{\cdot}10^{6}$	$9{\cdot}10^{-4}$	1.2–1.7
K15	73.7*	$3{\cdot}10^{4}$	$2{\cdot}10^{-3}$	30	4.89 ± 2.99	$5{\cdot}10^{5}$	$1{\cdot}10^{-3}$	0.8–4

https://doi.org/10.1371/journal.ppat.1011368.t001

Table 2. Data collection and refinement statistics.

	CagI:K2	CagI:K5
Wavelength	0.9786	0.9786
Resolution range	45.03–1.836 (1.901–1.836)	43.88–2.001 (2.073–2.001)
Space group	P 21 21 21	P 21 21 21
Unit cell	32.441 79.673 90.056 90 90 90	32.47 79.769 87.755 90 90 90
Total reflections	281017 (26889)	141192 (13436)
Unique reflections	21125 (2038)	15862 (1493)
Multiplicity	13.3 (13.1)	8.9 (9.0)
Completeness (%)	99.59 (97.46)	98.60 (95.27)
Mean I/sigma(I)	15.79 (1.79)	12.33 (1.81)
Wilson B-factor	23.01	25.75
R-merge	0.2503 (1.337)	0.1754 (1.123)
R-meas	0.2603 (1.39)	0.1861 (1.19)
R-pim	0.07087 (0.3776)	0.06135 (0.3907)
CC1/2	0.998 (0.682)	0.995 (0.657)
CC*	1 (0.901)	0.999 (0.891)
Reflections used in refinement	21071 (2036)	15832 (1491)
Reflections used for R-free	1015 (95)	783 (58)
R-work	0.1616 (0.2487)	0.1782 (0.2525)
R-free	0.1941 (0.2750)	0.2144 (0.2594)
CC(work)	0.970 (0.869)	0.958 (0.845)
CC(free)	0.948 (0.852)	0.938 (0.644)
Number of non-hydrogen atoms	1895	1847
macromolecules	1689	1683
ligands	0	0
solvent	206	164
Protein residues	224	223
RMS(bonds)	0.007	0.011
RMS(angles)	1.07	1.36
Ramachandran favoured (%)	99.55	98.63
Ramachandran allowed (%)	0.45	1.37
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.00	0.00
Clashscore	1.79	3.28
Average B-factor	23.22	26.48
macromolecules	21.81	25.77
solvent	34.84	33.83
Number of TLS groups	13	

https://doi.org/10.1371/journal.ppat.1011368.t002

(from the CagI:K5 complex) with AF model monomer showed that the prediction was indeed remarkably correct, with a rmsd of 0.9 Å for 103 Cα (Fig 3B). The K5 and K2 interaction site of CagI consists of a hydrophobic groove formed by α4 and α5 residues. Interactions between DARPins and CagI are widespread along the groove and extend to the concave face formed by the DARPin variable loops that wrap around α5 (Fig 3C). While the interface relies mostly on hydrophobic interactions, five hydrogen bonds also exist between CagI and DARPin residues, respectively: T235—D112, S242—D79, S242—H50, E210—Q28 and S254—E22 (Fig 3C). A single residue difference between K2 and K5 in the C-cap moiety of the DARPin generates small but significant changes in the DARPin/CagI interface. At position 125 the, K5 residue is a

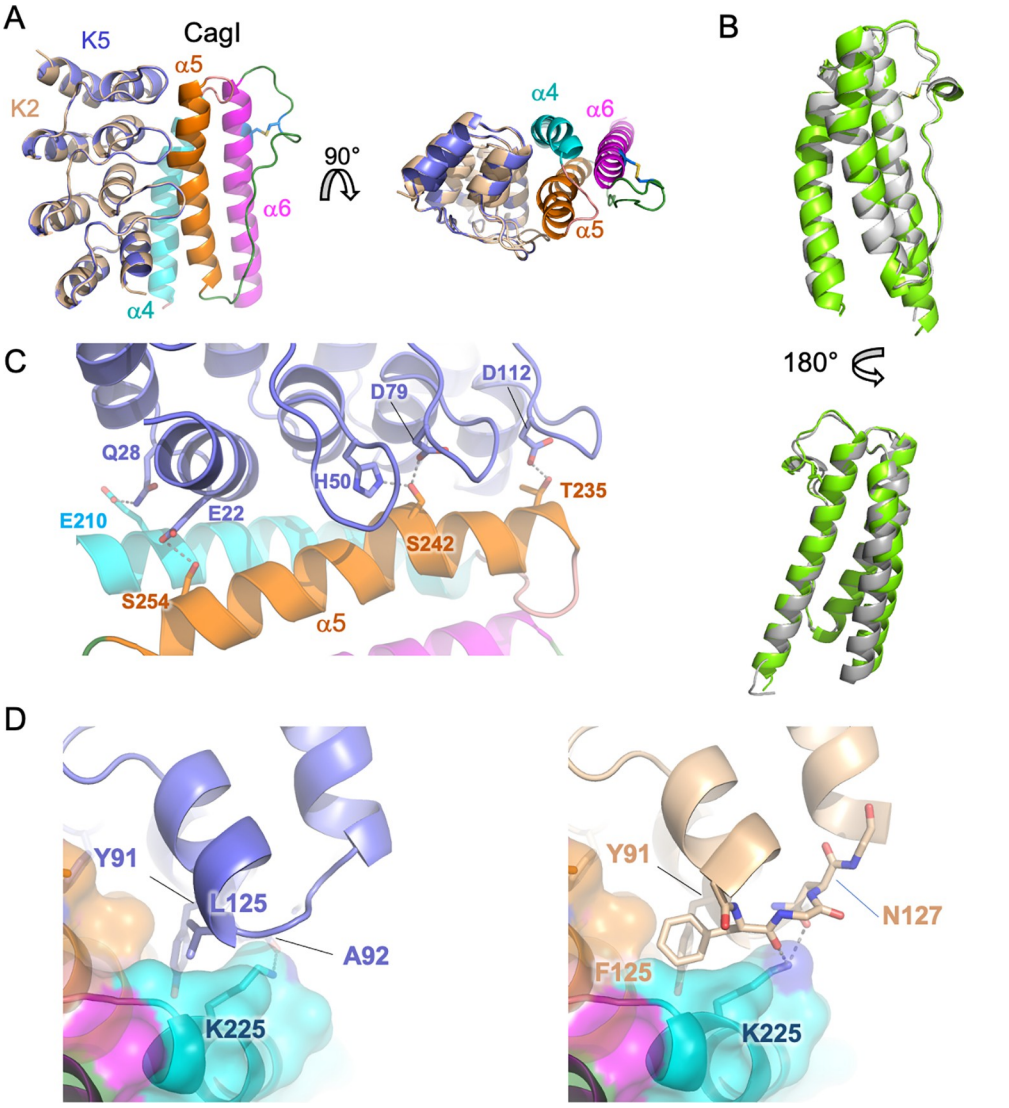


Fig 3. Structures of CagI:DARPin complexes. A) Overview of the structure of CagI:K2 and CagI:K5 complexes. The two structures of DARPin K2 (wheat) and K5 (slate) have been superimposed and are depicted as cartoons. The CagI molecule is displayed as cartoon and surface coloured according to secondary structure (α4 in cyan, α5 in orange and α6 in magenta). Side chains of cysteines 272 and 283 involved in disulfide bridges are shown as ball-and-stick with atoms coloured blue (carbons) and yellow (sulfur). For clarity, only the CagI molecule from the CagI:K2 complex is displayed. B) Two rotated views of the structural superimposition of the crystal structure of CagI$^{204-307}$ from the CagI:K2 complex (green) and corresponding region of the AF model (grey) displayed as cartoon. The side chains of the cysteine residues 272 and 283 forming the disulfide bond are displayed as ball-and-stick with sulfur atoms coloured in yellow. C) Detailed view of the interface between DARPin K5 loops and CagI α5 with side chains involved in hydrogen bonds shown as ball-and-stick with atoms coloured as follows: nitrogen in blue, oxygen in red, carbon coloured as in A). Grey dashed lines indicate hydrogen bonds. D) Close-up view of the interface of CagI:K5 (left panel) and CagI:K2 (right panel) showing residues F125 in K2 and L125 in K5 binding to the CagI groove.

https://doi.org/10.1371/journal.ppat.1011368.g003

leucine and its side chain is placed near a pocket formed by CagI α5 residues T235 and L239 and α4 residues L228 and K225 (Fig 3D). In K2, the phenylalanine 125 inserts deeper into the pocket and is stabilised by a T-shaped pi-stacking with the nearby Y91. In addition, the side chain of CagI residue K225 makes two hydrogen bonds with the carbonyl of the F125 main chain and the N127 side chain of DARPin K2 (Fig 3D). As a consequence, the surface buried by complex formation is slightly increased in CagI:K2 with 994 Å^2 compared to 974 Å^2 in CagI:K5.

Statistics for the highest-resolution shell are shown in parentheses.

DARPins K2 and K5 interact with two CagI molecules

Analysis of the crystal packing indicates a second interaction site in the CagI:K2 and CagI:K5 structures with significant scores in PISA [36], burying around 800 Å^2. In addition to the asymmetric unit CagI, K2 and K5 interact with a symmetry-related CagI fragment (noted CagI'). The interface named region 2 (region 1 being the first interface described above) covers CagI' α4 and part of α6 and involves the convex face of the three variables loops of the DARPins (Fig 4A). The interactions in that region are the same in CagI:K2 and CagI:K5 complexes. The interface relies on hydrophobic interactions at the groove formed by α4 and α6 and several hydrogen bonds and salt bridges between DARPins D74 and CagI K206 and K213 (Fig 4B). We therefore evaluated by SEC-MALS the stoichiometry of all CagI/DARPin complexes. CagI:K2, CagI:K5, CagI:K10, and CagI:K8 had a mass of 93–95 kDa, and CagI:K11 had a mass of 90 kDa, all consistent with a monomer of DARPin in complex with a dimer of CagI (Figs 4C and S7). CagI:K9 and CagI:K12 calculated masses were 117 kDa and 123 kDa, respectively, consistent with 2 DARPins:2 CagI complexes. No mass could be calculated with CagI:K15 given that the complex was aggregated. Thus we concluded that K2, K5, K8, K10 and K11 were able to bind simultaneously to two CagI^C molecules while K9 and K12 bind a single CagI^C.

DARPins targeting CagI inhibit CagA translocation in human adenocarcinoma gastric cells

To determine if the DARPins have an effect on CagA translocation by the *cag*T4SS, we next used a β-lactamase (TEM-1) assay as a reporter system for CagA type IV secretion into gastric

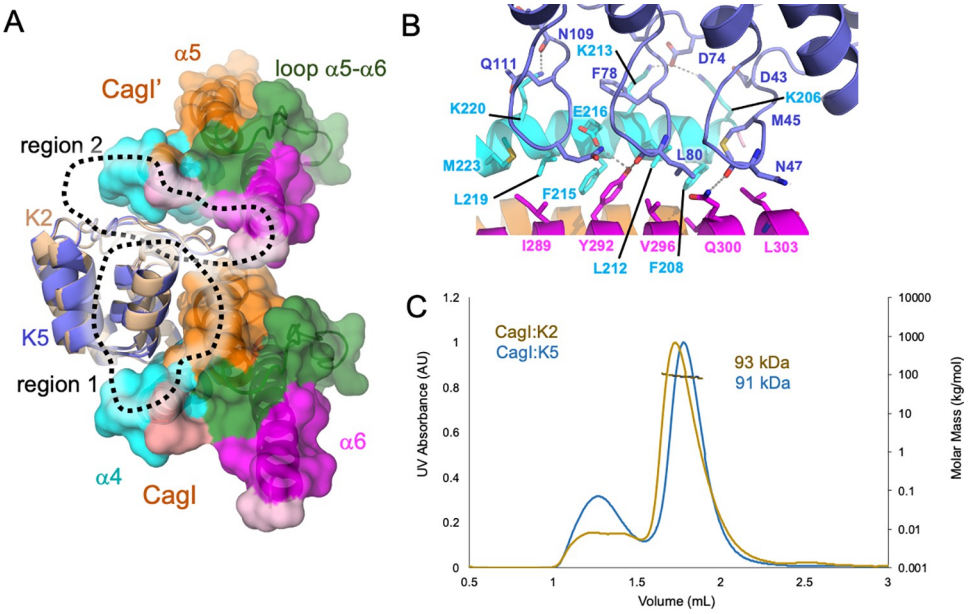


Fig 4. Structural basis for higher affinity of DARPin K2 on CagI. A) Overview of the structures of the CagI:K2 and CagI:K5 complexes interfaces. The two DARPin structures K2 (wheat) and K5 (blue) have been superimposed and are depicted as cartoons. The CagI molecule and symmetry related CagI' are displayed as cartoons and surfaces are coloured as in Fig 3. For clarity, only CagI molecules from the CagI:K2 complex are displayed. B) detailed view of region 2 interactions between DARPin K5 loop residues and CagI α4 and α6 with involved side chains displayed as ball and sticks with atoms coloured as follows: nitrogen in blue, oxygen in red, carbon as in A). Dashed lines indicate hydrogen bonds. C) SEC-MALS measurements of the CagI:K2 and CagI:K5 purified complexes.

https://doi.org/10.1371/journal.ppat.1011368.g004

epithelial cells [37]. Bacteria producing a TEM-1-CagA fusion were co-incubated with a gastric adenocarcinoma cell line (AGS) in the presence or absence of DARPins, and translocation of TEM-1-CagA was determined via blue-to-green fluorescence ratios after loading the cells with CCF4-AM. We observed that bacteria incubated with DARPins K9, K10, K11, K12 and K15 translocated CagA as efficiently as the untreated P12[TEM-1-CagA] (Fig 5A). In contrast, a

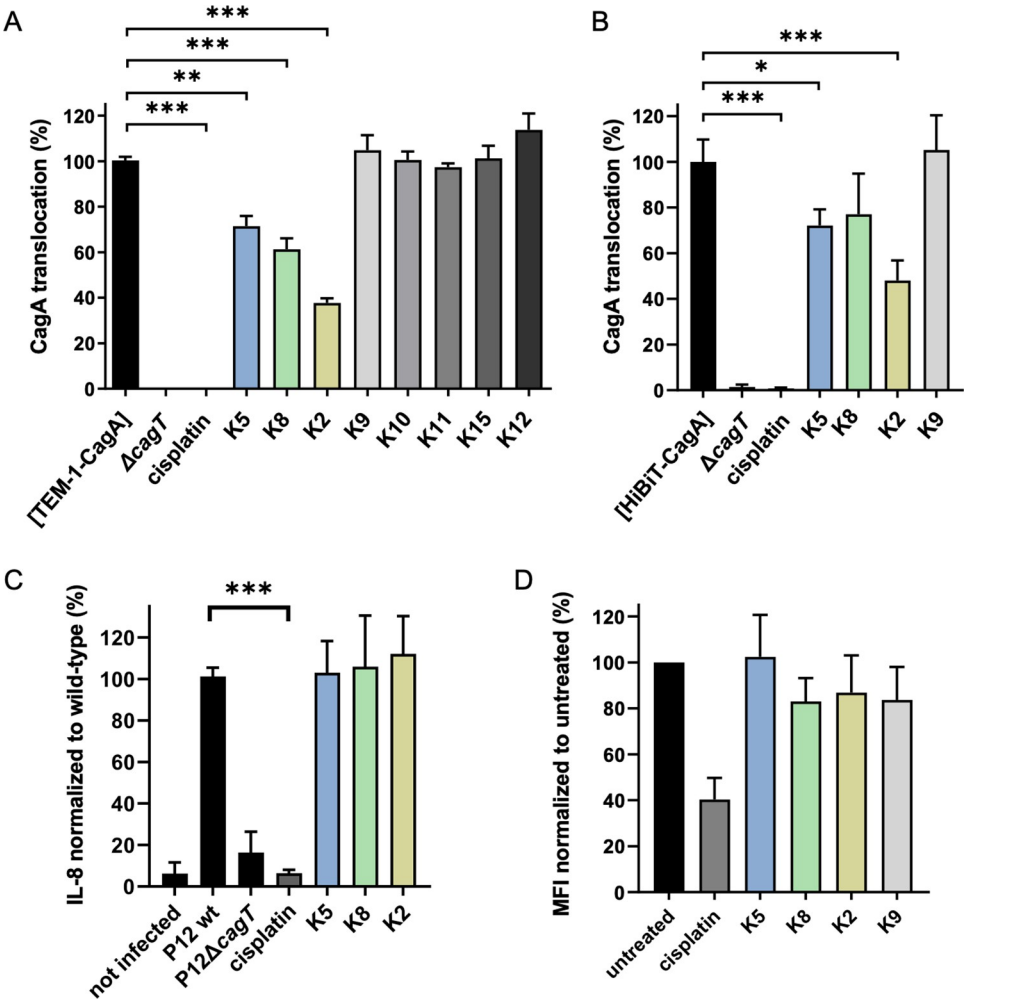


Fig 5. CagA translocation inhibition by DARPins. A) *H. pylori* P12 [TEM-1-CagA] was co-incubated for 2.5 h with AGS cells in the absence or presence of the indicated DARPins at a concentration of 5 μM, and CagA translocation was determined by a TEM-1-CagA translocation assay. As controls, the secretion-deficient mutant P12Δ*cagT* [TEM-1-CagA] was used without pre-treatment, or P12 [TEM-1-CagA] was pre-incubated for 30 min with 100 μM cisplatin. Data are indicated in relation to untreated control, which was set to 100%, and they represent mean values and standard deviations of five independent experiments. (One-way ANOVA; Tukey post-hoc test; **, $p<0.01$; ***, $p<0.001$). B) *H. pylori* P12 [HiBiT-CagA] was either left untreated or pre-treated for 30 min with 5 μM of the indicated DARPins or 100 μM cisplatin in PBS/10% FCS at 37˚C, 10% CO_2, and the bacterial suspensions were used to infect AGS [LgBiT] cells for 2.5 h. Luminescence values were recorded and normalized to untreated control. Data are indicated as mean values with standard deviations resulting from four independent experiments. (One-way ANOVA; Tukey post-hoc test; *, $p<0.05$; ***, $p<0.001$). C) IL-8 levels were determined in AGS cell supernatants 4h after infection with P12 wild-type (wt), P12Δ*cagT* or P12 wt in the presence of 100 μM cisplatin, K5, K8 or K2 DARPins (5 μM). IL-8 concentrations were normalized to untreated P12 wt control, and are indicated as mean values and standard deviations of six independent experiments. (One-way ANOVA; Tukey post-hoc test; ***, $p<0.001$). D) Cell binding measurements by flow cytometry with *H. pylori* P12 [pHel12::gfp] pre-treated with DARPins at 5 μM, cisplatin at 100 μM, or left untreated. Data shown are mean values and standard deviations of median fluorescence intensities (MFI) normalized to untreated bacteria, resulting from three independent experiments.

https://doi.org/10.1371/journal.ppat.1011368.g005

30% reduction was seen with bacteria incubated with K5, and even a reduction by 40% and 60% in CagA translocation after pre-incubation with K8 and K2, respectively (Fig 5A). In contrast, the potent, but unspecific type IV secretion-disrupting agent cisplatin/DMSO [38], inhibited CagA translocation completely. Similar results were obtained with a HiBiT-CagA translocation assay, which uses a split-luciferase system as a translocation reporter [39]. In this assay, pre-treatment of P12[HiBiT-CagA] with DARPins K5, K8 and K2 also resulted in a decrease of translocation to comparable levels as in the TEM-1-CagA translocation assay (even though K8 did not reach statistical significance), whereas treatment with K9 did not decrease translocation efficiency (Fig 5B). Another hall-mark of *cag*T4SS-activity is the increase in IL-8 production by the host cell. We thus measured IL-8 induction in AGS cells in the presence or absence of DARPins (Fig 5C). No difference could be observed in IL-8 induction demonstrating that the DARPins K2, K5 and K8 had no effect on this phenotype. To determine if the CagA translocation inhibition was due to a lack of attachment of the whole bacteria to the AGS cells, we monitored bacterial adhesion via flow cytometry using *H. pylori* cells expressing GFP. The results show no major differences in cell adhesion upon DARPin treatment, whereas cisplatin resulted in a strong reduction in bacterial adherence (Fig 5D), suggesting that the same DARPins did not affect *H. pylori* binding to AGS cells. Finally the DARPins K2, K5, K8 and K9 had no effect on viability of adherent *H. pylori* after incubation, while no bacteria survived the cisplatin treatment (S8 Fig).

CagI mediates cell binding and spreading via its C-terminal domain

Previous studies showed that CagL mediates AGS cell adhesion *in vitro* in a manner reminiscent of fibronectin [20]. Given that CagI and CagL share structural homology, we investigated if CagI could also have a similar effect on human cells. An assay was implemented to compare the binding of AGS cells to purified CagL, CagI or CagI N and C-terminal domains. Multiwell plates were first coated with different amounts of the proteins and then incubated with AGS cells. The extent of AGS cell adhesion was measured 60 min after cell seeding using a colorimetric reaction as described in Materials and Methods. While $CagI^{N}$ did not induce any cellular adhesion, AGS cells adhered to the three other substrates in variable proportions. CagL and $CagI^{C}$ induced the strongest cellular adhesion (Fig 6A). A variation of the cellular morphology could be noticed according to the ligands (Fig 6B).While the majority of the cells remain rounded or slightly spread out after both CagI and CagL binding, all the cells appear fully spread on $CagI^{C}$, suggesting a rapid recruitment and organization of the actin cytoskeleton after adhesion. There, AGS cells appear fusiform, projecting cytoplasmic extensions indicating rearrangement similar to those induced by extracellular matrix proteins such as fibronectin (Fig 6B). Cell spreading induced by $CagI^{C}$ was significantly higher than CagI or CagL as seen by measurements of cell surface, cell perimeters or cell Feret's diameter (S9 Fig). Next, we determined if K2 was able to inhibit CagI and $CagI^{C}$ binding to AGS cells. After coating with the target proteins, the plates were incubated with increasing amount of DARPins prior to incubation with AGS cells. As seen in Fig 6C, DARPin K2 efficiently inhibits AGS cell binding to full-length CagI or $CagI^{C}$ in a dose-dependent manner but had no effect on AGS cell binding to CagL. Inhibition was also observed with K8 on cell adhesion to CagI but not on $CagI^{C}$ (S10 Fig). In similar conditions, K11showed a weak inhibition on cell adhesion on CagI (Fig 6D). However, the inhibition of K11 was not dose-dependent, and no effect was observed on $CagI^{C}$ (Fig 6D). K10 and K12 did not show any effect on AGS cells binding (S10 Fig). This suggests that the CagI epitopes targeted by K2, and to a lesser extent K8 but not K10, K11 and K12 are important for cell binding.

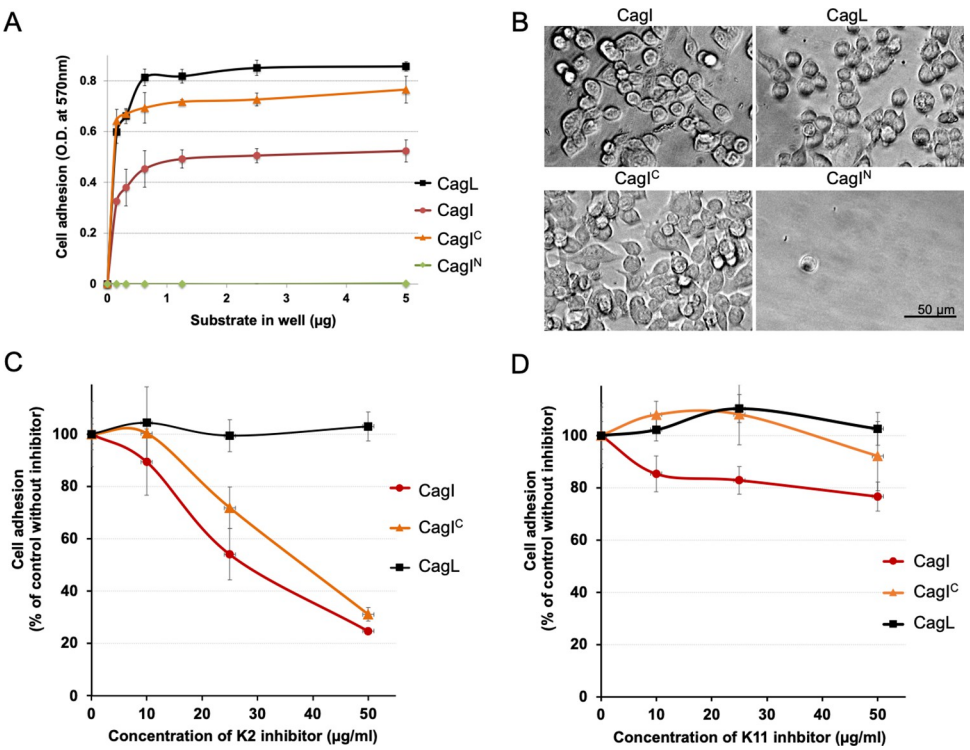


Fig 6. Cell binding and spreading to *cag*T4SS proteins and domains. A) Dose-dependent AGS cell adhesion to CagL, CagI, CagI^{C} and CagI^{N}. Multiwell plates were coated with different amounts of the proteins as indicated on the Figure. Each assay point was derived from triplicate measurements. B) Representative images of adhered AGS cells on well-surfaces coated with 0.15 μg of indicated proteins. C) Effect of DARPin K2 or D) K11 on adhesion of AGS cells to CagL, CagI, CagI^{C}. Multiwell plates were coated with 4 μg of each protein. After saturation with 1% BSA, the wells were incubated with 50 μL of the indicated concentration of K2 or K11 for 1 h at room temperature, and the cells were seeded in the presence of the inhibitor. The extent of adhesion was measured as previously described and expressed as percentage of adhesion to each protein in the absence of the inhibitor. Each assay point was derived from triplicate measurements.

https://doi.org/10.1371/journal.ppat.1011368.g006

Discussion

*H. pylori cag*T4SS encodes for an unusually large number of proteins able to mediate host cell interaction, probably illustrating the long co-evolution of the bacteria with its host [40,41]. Thanks to a remarkable genomic plasticity, the bacteria has evolved a plethora of mechanisms to adapt to human populations, and even different niches within an infected stomach [42]. While the CagA injection mechanism is still poorly understood, the *cag*T4SS pilus is essential to interact with the host cell and to deliver the oncoprotein [18]. Proteins associated with the *cag*T4SS pilus production might have different functions. CagL not only interacts with integrins but also with the TLR5 receptor and this is also the case for CagY repeat region II [43,44]. Structural and functional information is available on CagL [19,24,25], CagA [45,46] and CagY [4,11,47], but little is known about the CagI protein despite its essential role in pilus formation [16].

Using an integrative structural biology approach, we show here that CagI forms an elongated dimer assembled via interactions between N-terminal domains followed by two C-terminal domains that are monomeric by themselves. On the one hand, the association mode of the N-terminal part was poorly predicted by AF, and although SAXS data confirms its general architecture, additional studies will be required to obtain structural details on this portion of

CagI. On the other hand, our study, along with previous sequence analysis [16,25], establish that CagL and the CagI C-terminal domain share structural similarity, including a set of three helices and a disulfide bridge. The resemblance between the two proteins extends to their localization: both were found present in the periplasm, and surface-exposed [30,31]. CagL localizes at the tip of the pilus [21] but this is not entirely clear for CagI, although immunolocalization and electron microscopy identified CagI in *cag*T4SS pilus-like appendages [30]. Our work shows that similarly to CagL, $CagI^{C}$ is able to mediate specific interactions with the host cell. Given that DARPins reducing this interaction also inhibit CagA injection, we assume that surface-exposed CagI interactions with host cell components are important for delivery through the *cag*T4SS.

Since we found that CagI and CagL interact *in vitro* [28] and *in vivo* [31] it is tempting to speculate that both CagL and CagI could be present at the surface or on the *cag*T4SS pilus where they could mediate adhesion and CagA translocation. Along these lines, both CagL and CagI were able to interact with integrin $\alpha_5\beta_1$ [28]. Here we show that CagI can mediate AGS cell attachment and spreading in a similar manner as CagL. However, the two proteins are unlikely to use the same determinant to interact with the host cell. CagL was proposed to interact with integrin via its RGD motif (reviewed in [18]) but this feature and surrounding motifs [18], are absent in CagI (S3 Fig). The CagL D1 motif involved in TLR5 recognition [44] is also not conserved in CagI (S3 Fig). Conversely, data presented in our study suggest that the motifs of CagI targeted by the K2 and K5 DARPins are important for host cell interactions, and these are also not conserved in CagL. The function of CagL and CagI are thus not redundant and instead might complement each other during *H. pylori cag*T4SS interaction with the host cell.

That CagI and CagL are surface associated proteins involved in host cell recognition is reminiscent of what is described in some other T4SSs. In *Agrobacterium tumefaciens*, VirB5 is located at the tip of the VirB/D T4SS pilus and also promotes T-DNA translocation when added externally [48]. In the conjugative T4SSs from the plasmid pKM101, the VirB5 homologue TraC is also involved in host cell recognition and pilus adhesion [49]. TraC also interact with a protein named Pep (for PRD1 entry protein) with which it forms cluster at the surface of the bacterium to stimulate cell-to-cell contacts [50]. In the case of *H. pylori*, CagI and CagL, being exposed at the surface, might play a role in host-cell recognition. Indeed, while both proteins interact with integrins, the presence of these receptors is not essential for delivery of CagA from a mechanistic point of view [51]. However, *H. pylori* cells produce *cag*T4SS pilus and inject CagA at the basolateral surface, where integrins are located, but not at the apical sides of the cells [52]. Thus, CagI and CagL might sense the presence of integrins to trigger pilus assembly. Alternatively, CagI and CagL might be part of a larger, surface-accessible complex such as the lateral pores that have been observed on the *cag*T4SS pilus [14].

The results presented here identify a novel and effective target to prevent CagA delivery into host cell, hereby disarming the main oncogenic factor of *H. pylori*. Therapeutic strategies that target key virulence factors of pathogenic bacteria, while not actually killing the cells themselves, could prove to be vital for the treatment of numerous diseases [53]. In this regard, extracellular, species-specific appendages are appealing targets to prevent bacterial adhesion or effector translocation. Not only are they essential for virulence, but they have also the advantage of being easier to target, being accessible from the external milieu. Targeting pilus-associated proteins by small molecules or antibodies has been successful to disarm type III secretion systems [54]. In the case of *H. pylori cag*T4SS, efforts have so far been directed towards the screening or design of small molecules or peptides to target cytoplasmic components [55,56]. Some inhibitors have shown some efficiency at inhibiting the cytoplasmic VirB11 ATPase [57,58]. Other compounds inhibit pilus biogenesis and/or CagA translocation, but their mode of action remains undetermined [59]. The DARPin inhibition described here is a first example

of effector translocation inhibition by a small protein binder targeting the external part of the *cag*T4SS.The exact mechanism of this inhibition remains to be established but, interestingly the DARPins did not inhibit IL-8 production by the gastric cells. This suggests that the *cag*T4SS remains partly functional and thus able to translocate other pro-inflammatory molecules. IL-8 production is induced by heptose metabolites or peptidoglycan that are actively translocated by the *cag*T4SS (reviewed in [60]). This also suggests that the DARPins do not affect the assembly of the apparatus itself, which might also explain why the inhibition did not exceed 65%. In our AGS cell binding assays, K2 did not fully inhibit CagI or $CagI^{C}$ binding perhaps reflecting that some binding to the host cell might still occur and hence allow for CagA translocation. Moreover, despite the DARPin binding to CagI, other *cag*T4SS proteins interacting with host cell receptors, such as CagY or CagL, might allow the delivery of CagA. Although the *cag*T4SS inhibition is far from being complete, our study paves the way for strategies targeting *H. pylori*-specific extracellular determinants of CagA injection.

Material and methods

Protein structure prediction

We used the ColabFold Notebook for accessing AlphaFold2 Multimer [34,61] to submit the CagI sequence (strain 26695 Genebank: AAD07606) to structure prediction for three models of a dimer. Other sequences were also used to evaluate the variability of CagI structures. Figures were generated with Pymol (Schrödinger) using the output PDB files containing the per residues LDDT scores.

DNA manipulation, cloning, expression and protein purification

CagI, $CagI_{avi}$ and $CagI_{strep}$. Untagged CagI protein (residues 21 to 381) was purified as described in [28] using the pRSFMBP-*cagI* vector and *E. coli* BL21 (DE3) cells with the exception that detergent was removed by using an additional purification step. After HisMBP-tag cleavage CagI was loaded on a SOURCE 15Q 4.6/100 column (Cytiva) in a buffer containing 50 mM Tris pH 8, 20 mM NaCl and 0.005% DDM and washed with buffer 50 mM Tris pH 8, 20 mM NaCl with 10 column volumes. The protein was eluted with a liner gradient of buffer (50 mM Tris pH 8, 1M NaCl). The protein was then purified by size exclusion chromatography on a S200 Superdex column in 50 mM Tris pH 8, 150 mM NaCl.

The expression vector for $CagI_{avi}$ was generated by insertion of an avitag (GLNDIFEAQKIEWHE) at the 3' end of the *cagI* sequence in pRSFMBP-*cagI*. The $CagI_{avi}$ protein was purified as above and biotinylated with the BirA enzyme as described in [62].

For $CagI_{Strep}$ construction, expression and purification, the sequence corresponding to CagI residues 21–381 (from strain 26695) were fused to a C-terminal glycine-linker strep-tag encoding sequence (5' GGTGGAGGTTCTGGCGGTGGATCGGGAGGTTCAGCGTGGTC TCATCCTCAATTTGAAAAA 3'). $CagI_{Strep}$ was co-expressed with $CagL_{His}$ protein (residues 21 to 237 fused to c-terminal his_6-tag encoding sequence) in a pRSF duet vector. BL21 (DE3) cells carrying pRSF-$CagI_{Strep}$-$CagL_{His}$ expression vector were grown in LB at 37˚C until an OD_{600} of 0.6–0.8. Protein expression was induced for 16 h at 20˚C after adding 0.5 mM IPTG. For purification, cells were resuspended in lysis buffer (50 mM Tris pH 7.4, 200 mM NaCl) supplemented with protease inhibitor tablets (complete EDTA-free; Roche, one tablet per 250 mL of lysis buffer), lysozyme (0.1 mg/mL, Sigma-Aldrich) and DNase I (20 μg/mL, Sigma-Aldrich), and disrupted with three passages through a cell disrupter system (Constant Systems) operating at ~15,000 psi. The fraction containing $CagI_{Strep}$ was separated from the soluble fraction after centrifugation at 7000 g for 20 min. The pellet was then resuspended in 50 mM Tris pH 7.4, 200 mM NaCl using a Dounce homogenizer and solubilized by addition of DDM to a

final concentration of 0.5% by stirring at medium speed for 16 hours at 4˚C. Insoluble material was pelleted by ultracentrifugation at 200,000 g for 1 hour at 4˚C and the supernatant was loaded onto a StrepTrap column (Cytiva). The column was washed with 50 mM Tris pH 8, 150 mM NaCl, 0.005% DDM and $CagI_{Strep}$ was eluted with 50 mM Tris pH 8, 150 mM NaCl, 0.005% DDM supplemented with 2.5 mM desthiobiotin. The elution fraction was analysed by SDS-PAGE before to loading on a HisTrap column (Cytiva) to remove $CagL_{His}$ proteins. The flow-through, containing $CagI_{Strep}$ was collected, concentrated and loaded onto a Superdex 200 HiLoad 16/600 gel filtration column (Cytiva) equilibrated in 50 mM Tris pH 8, 150 mM NaCl and 0.005% DDM.

CagI domains. PCR fragments encoding for CagI residues 27–190 ($CagI^N$) or residues 191 to 381 ($CagI^C$) were amplified using primers *cagI*^{N}*fw* (5'-CACCACGCTTGAACCCGCC TTAAAAG-3'), *cagI*^{N}*rev* (5'- TCAACTTCCTAGAGCTTGAGAAAG), *cagI*^{C}*fw* (CACCTCTT CTGACAACGCTCAATACATC), *cagI*^{C}*rev* (TCATTTGACAATAACTTTAGAGCTAG) and inserted into the pET151D topo vector (ThermoScientific) following the manufacturer's protocol. The resulting vectors pET151$CagI^N$ or pET151$CagI^C$ encode each CagI domain fused to a N-terminal His_6-tag followed by a tobacco-etch virus (TEV) cleavage site. *E. coli* T7 Express cells harboring pET151$CagI^N$ or pET151$CagI^C$ were grown in 1 L LB medium supplemented with ampicillin (100 μg/ml) at 37˚C until OD_{600nm} reached 0.6. Protein expression was induced for 16 hours at 20˚ C by adding 1 mM IPTG (final concentration). Cells were harvested and resuspended in 20 ml of buffer AG (50 mM Tris pH 8, 200 mM NaCl, 5% glycerol (v/v)). Solutions containing the cells were supplemented with Triton-X100 (final concentration 1%), one tablet (per 250 ml of buffer) of complete EDTA-free protease inhibitor (Roche), lysozyme (0.1 mg/ml, Sigma-Aldrich) and DNase I (20 μg/ml, Sigma-Aldrich) prior to sonication. Cell debris was removed by centrifugation (14,000 g, 4˚ C, 15 min) and the supernatants were loaded onto a HisTrap column (Cytiva). Proteins were eluted with a 0 to 100% linear gradient of buffer AG containing 500 mM imidazole. Fractions containing CagI domains were pooled and His_6 tags were cleaved by TEV protease with 0.5 mM EDTA and 5 mM DTT and dialysed 16 h against buffer AG at 4˚C. Proteins were loaded onto HisTrap column and flow-through containing cleaved protein was pooled and concentrated (Amicon 3 K Sigma Aldrich). Proteins were loaded onto a Superdex 200 increase 10/300 GL column (Cytiva) equilibrated in buffer AG.

Selection and screening of DARPins

To generate DARPin binders, CagI biotinylated at a C-terminal avi tag was immobilized, in alternating selection rounds, on either MyOne T1 streptavidin-coated beads (Pierce) or Sera-Mag neutravidin-coated beads (GE). Ribosome display selections were performed essentially as described [37], using a semi-automatic KingFisher Flex MTP96 well platform.

The library includes N3C-DARPins with the original randomization strategy as reported [63] but also a stabilized C-cap [33,64,65]. Additionally, the library is a mixture of DARPins with randomized and non-randomized N- and C- terminal caps, respectively [33,66]. Successively enriched pools were ligated as intermediates in a ribosome display-specific vector [66]. Selections were performed over four rounds with decreasing target concentration and increasing washing steps, and the third round included a competition with non-biotinylated CagI to enrich for binders with high affinities.

The final enriched pool of cDNA encoding putative DARPin binders was cloned as fusion construct into a bacterial pQE30 derivative vector (Qiagen), containing a T5 lac promoter and lacIq for expression control, with a N-terminal $MRGS(H)_8$ tag and C-terminal FLAG tag via unique *Bam*HI and *Hind*III sites. After transformation of *E. coli* XL1-blue, 380 single DARPin

clones selected to bind CagI were expressed in 96-well format and lysed by addition of B-Per Direct detergent plus lysozyme and nuclease (Pierce). After centrifugation these crude extracts were used for initial screening to bind CagI using ELISA. For IMAC purification of DARPins they were expressed in deep-well 96-well plates, lysed with Cell-Lytic B (SIGMA) and purified over a 96-well IMAC column (HisPur Cobalt plates, Thermo Scientific).

ELISAs were performed using streptactin-coated plates (iba-lifesciences) and used for immobilization of $CagI_{strep}$ at a concentration of 50 nM. Detection of DARPins (1:1000 dilution of crude extracts, or a concentration of 50 nM for IMAC-purified DARPins) binding to CagI was performed using a mouse-anti-FLAG M2 monoclonal antibody (dilution 1:5000; Sigma, F1804) as primary and a goat-anti-mouse antibody conjugated to an alkaline phosphatase (dilution 1:10,000; Sigma, A3562) as secondary antibody. After addition of pNPP (para-nitrophenyl phosphate) absorbance at 405 nm was determined after 30 minutes. Signals at 540 nm were subtracted as background correction.

DARPin purification and pull downs assays

pQE30 expression vectors expressing the DARPins K5, K9, K2, K12, K15, K8, K10 and K11 (described above) were transformed into *E. coli* T7 Express cells (NEB). Protein expression and extraction were performed as described above for CagI domains in buffer A (50 mM Tris pH 8, 150 mM NaCl) for K5 or buffer AG (50 mM Tris pH 8, 150 mM NaCl, glycerol 5% (v/v)) for others DARPins. Supernatants from 1L bacterial cell cultures were loaded onto a HisTrap column (Cytiva), washed successively with 3 column volumes of buffer AG (or buffer A for K5) 2 column volumes of buffer AG (or A for K5) supplemented with 1 M NaCl. Proteins were eluted with a 0 to 100% linear gradient of buffer AG (or buffer A for K5) containing 500 mM imidazole. Fractions containing the DARPins were pooled, concentrated (Amicon 3 K Sigma Aldrich) and loaded onto a Superdex 200 increase 10/300 GL column (GE healthcare) equilibrated in buffer AG or buffer A (for K5).

$CagI_{Strep,}$ $CagI^{N}$ and $CagI^{C}$ were mixed and incubated on ice for 1 h with 50 μg of $_{His8}$DARPins at a 2-fold molar excess of CagI or CagI domains. The mixtures were incubated with 30 μL Ni-NTA magnetic beads (Merck) and loaded onto Biosprint 15 (Qiagen) for purification. Proteins were washed twice with 750 μL of buffer AG or A (for CagI:K5) and finally eluted in 150 μL buffer A (CagI:K5) or AG containing 500 mM imidazole.

Co-expression of DARPins and $CagI_{strep}$

pRSF-$cagI_{strep}$ and pQE30-DARPin vectors were introduced in *E. coli* T7 Express cells (NEB). Cells were grown in 50 mL LB medium supplemented with kanamycin (50 μg/mL) and ampicillin (100 μg/mL) at 37˚C until OD_{600nm} reached 0.6. Protein expression was induced for 16 hours at 20˚ C by adding 1 mM IPTG (final concentration). Cells were harvested and resuspended in 1 ml of buffer A (50 mM Tris pH 8, 150 mM NaCl) for CagI: K5 or buffer AG for the remaining CagI:DARPin complexes. For cell lysis buffers were supplemented with Triton-X100 (final concentration 1%), one tablet (per 250 mL of buffer) of complete EDTA-free protease inhibitor (Roche), lysozyme (0.1 mg/mL, Sigma-Aldrich) and DNase I (20 μg/mL, Sigma-Aldrich). The cells were sonicated and centrifuged at 14,000 g 4˚ C for 15 min. Supernatants were incubated with 30 μL nickel magnetic beads (Merck) and loaded onto Biosprint 15 (Qiagen). Proteins were washed two times with 750 μl of buffer A (or AG) and eluted with 150 μL buffer A (or AG) supplemented with 500 mM imidazole. For western-blot detection, proteins samples were separated on a 20% SDS-PAGE, transferred to a polyvinylidene difluoride membrane, and immunoblotted using a mouse monoclonal antibody against Strep-tag (Qiagen, 34850). Alkaline phosphatase conjugated to anti-mouse IgG (Sigma-Aldrich, A1293) was used

as a secondary antibody. Detection was performed by colorimetry using nitro blue tetrazolium chloride / 5-bromo-4-chloro-3-indolyl-phosphate (NBT-BCIP, Sigma-Aldrich) as a substrate.

For large scale purification, CagI:DARPins expression and extraction of proteins were performed as described above except for a larger volume of cell culture (1 L). After lysis by sonication and centrifugation (14,000 g, 4˚C 15 minutes), the CagI:DARPins supernatants were loaded onto a 5 mL StrepTrap column (Cytiva), washed with 3 column volumes of buffer AG (or A for CagI:K5) and eluted with buffer AG (or buffer A for the CagI:K5 complex) containing 2.5 mM desthiobiotin. Elution fractions were loaded onto a 5 mL HisTrap column (Cytiva) and after a 3 column volumes wash with buffer AG (or A for CagI:K5), proteins were eluted in buffer AG (or A, see above) containing 500 mM imidazole. Fractions containing the complexes were concentrated (Amicon 10 K Sigma Aldrich) and loaded onto a Superdex 200 increase 10/300 GL column (Cytiva) equilibrated in 50 mM Tris pH 8, 150 mM NaCl for CagI:K5 complex and 50 mM Tris pH 8, 200 mM NaCl, 5% glycerol v/v for other CagI:DARPins complexes.

Multi-angle light scattering (MALS)

Size-exclusion chromatography experiments coupled to multi-angle laser light scattering (MALS) and refractometry (RI) were performed on a Superdex S200 Increase 5/150 GL column with size-exclusion buffer 50 mM Tris pH 8, 150 mM NaCl for DARPin K5 and CagI:K5 complex and with buffer 50 mM Tris pH 8, 200 mM NaCl, 5% glycerol v/v for others CagI:DARPin complexes. Fifty microliters of proteins were injected at a concentration of 5 to 8 mg/mL. Online MALS detection was performed with a miniDAWN-TREOS detector (Wyatt Technology Corp., Santa Barbara, CA, USA) using a laser emitting at 690 nm and by refractive index measurement using an Optilab T-rEX system (Wyatt Technology Corp.). Weight-averaged molar masses (Mw) were calculated using the ASTRA software (Wyatt Technology Corp.).

Crystallization, structure determination and refinement

Crystals of CagI:K5 and CagI:K2 were obtained by the sitting drop vapor diffusion method using a Mosquito robot. Drops consisting of 200 nL of protein complex (7 mg/mL) with 200 nL of reservoir solution were left at 19˚C for two weeks. Crystals of CagI:K5 appeared in condition E5 of the PACT Premier screen (Molecular Dimension) with a reservoir solution consisting of 0.2 M sodium nitrate, 20% w/v PEG 3350. CagI:K2 crystals appeared in condition E2 of the same screen with a reservoir solution consisting of 0.2 M sodium bromide, 20% w/v PEG 3350. Crystals were flash frozen in reservoir solution supplemented with glycerol 15% (v/v). Data were collected at 100˚K at PROXIMA 1 beamline of the synchrotron SOLEIL and processed using XDS [67] and AIMLESS [68] from the CCP4 program suite [69,70]. Crystals of CagI:K5 and CagI:K2 diffracted to resolutions of 2.0 Å and 1.8 Å, respectively and belonged to the orthorhombic space group $P2_12_12_1$ with very similar cell dimensions (Table 2). The structure of CagI:K5 was solved by molecular replacement using the coordinates of DARPin E11 (PDB ID: 6FP8 [71]) as a probe in PHASER [72]. Examination of the resulting electron density indicated that additional helices were present in the asymmetric unit. Manual building resulted in a first model of several helices and the unit contained one K5 molecule and a fragment of CagI. After several rounds of manual building/refinement the sequence of the additional peptide could be undoubtedly attributed to CagI residues 204 to 307. The resulting model was used as a template to solve the structure of CagI:K2. The models were refined with final R_{work}/R_{free} of 0.18/0.23 (CagI:K5) and 0.17/0.20 (CagI:K2). The coordinates and structure factors were deposited in the Protein Data Bank with accession code 8AIW (CagI:K5) and 8AK1 (CagI:K2).

Small-angle X-ray scattering

SAXS data were collected for CagI and CagI domains at the ESRF BioSAXS beamline BM29 using an online size-exclusion chromatography setup. Fifty μl of protein (8 mg/mL) were injected into a size-exclusion column (S200 increase 5/150) equilibrated in 50 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol v/v. Images were acquired every second for the duration of the size-exclusion run. Buffer subtraction was performed by averaging 100 frames. Data reduction and analysis was performed using the BsxCuBE data collection software and the ATSAS package [73]. The program AutoGNOM was used to generate the pair distribution function ($P(r)$) and to determine D_{max} and R_g from the scattering curves ($I(q)$ *versus* q) in an automatic, unbiased manner. Theoretical curves from the models were generated by FoXS [74]. Ab initio modelling was performed with DAMMIN [75].

Surface plasmon resonance

Measurements were performed using a Biacore T200 instrument (Cytiva). CagI and CagI^C were covalently immobilised to the dextran matrix of a CM5 sensorchip via their primary amine groups. The carboxymethylated dextran surface was activated by the injection at 5 μL/min of a mixture of 200 mM EDC [*N*-ethyl-*N*′-(3-dimethylaminopropyl)carbodiimide] and 50 mM NHS (*N*-hydroxysuccinimide). Ligands were diluted in 10 mM sodium acetate pH 4 to a 10–20 μg/mL concentration before injection over the activated surface of the sensor chip. Residual active groups were blocked by injection of 1 M ethanolamine pH 8.5. Immobilization levels of 1,200 RU (response units) were obtained for CagI and 340 RU for CagI^C. A control flow cell was activated by the NHS/EDC mixture and deactivated by 1 M ethanolamine pH 8.5 without any coupled protein. Control sensorgrams were subtracted online from the sensorgrams to derive specific binding responses. Analytes were injected at 50 μg/mL for 120 seconds after dilution in running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 0.05% P20). The sensorchip surface was regenerated with 2 pulses (30 sec) of ethylene glycol 50% and 1 pulse of 2 M guanidine hydrochloride. The equilibrium K_D were calculated using the Biacore T200 evaluation software (v3.2.1).

Quantification of CagA translocation and bacterial cell binding

Translocation of CagA into AGS cells was determined quantitatively using either the TEM-1-CagA translocation assay [37], or the HiBiT-CagA translocation assay [39]. For the TEM-1-CagA assay, AGS cells were co-incubated with *H. pylori* P12 [TEM-1-CagA] for 2.5 h in 96-well microtiter plates in PBS/10% FCS. After infection, cells were loaded with fluorescent substrate CCF4-AM in a loading solution (LiveBLAzer-FRET B/G loading kit; Invitrogen) supplemented with 1 mM probenecid (Sigma) according to the manufacturer´s instructions. Cells were incubated with this loading solution at room temperature in the dark for 2 h, and then measured with a Clariostar reader (BMG Labtech) using an excitation wavelength of 405 nm, and emission wavelengths of 460 nm, or 530 nm. CagA translocation was calculated as the ratio of background-corrected emission values at 460 nm to 530 nm and normalized to *H. pylori* P12 [TEM-1-CagA] and P12Δ*cagT* [TEM-1-CagA] as positive and negative controls. For the HiBiT-CagA assay, *H. pylori* P12 [HiBiT-CagA] was pre-cultured in PBS/10% FCS for 2 h at 37˚C, 10% CO_2. Subsequently, AGS [LgBiT] cells seeded in a 96-well plate (4titude) were infected with 200 μl of this pre-culture, and incubated at 37˚C, 5% CO_2 for 2.5 h. Supernatants containing unbound bacteria were discarded, and cells were loaded with 40 μL PBS/FCS and 10 μL 5x luciferase substrate mix. After 10 min incubation, luminescence was measured at 470 nm in a Clariostar reader. The amount of translocated HiBiT-CagA was calculated after correction for the background signal as percentage in relation to the untreated *H. pylori* P12

[HiBiT-CagA] control. For inhibition experiments, bacteria were pre-incubated with the respective DARPins at a concentration of 5 μM for 30 minutes at 37˚C in PBS/10% FCS, followed by infection for 2.5 h in the presence of the DARPins.

For determination of bacterial binding to gastric cells, AGS cells were infected with *H. pylori* strain P12 [pHel12::gfp] [38], using an MOI of 60, and infection was allowed to proceed for 1 h at 37˚C and 5% CO_2. For inhibition experiments, bacteria were pre-incubated with the respective DARPins at a concentration of 5 μM, or cisplatin at 100 μM, for 30 minutes at 37˚C in PBS/10% FCS, and then co-incubated with AGS cells, as above. After three washing steps with PBS, cells with adherent bacteria were collected by EDTA treatment, and analysed for GFP fluorescence in a flow cytometer (FACS CantoII, BD Biosciences). For analysis, the median fluorescence intensity of non-infected cells was subtracted from that of infected samples.

To determine the viability of adherent bacteria, *H. pylori* P12 was either pre-incubated with DARPins at a concentration of 5 μM, or cisplatin at 100 μM, in PBS/10% FCS for 30 min at 37˚C, or left untreated under the same conditions, and then co-incubated at an MOI of 60 with AGS cells at 37˚C and 5% CO_2 for 2.5 h. Subsequently, supernatants containing unbound bacteria were removed, cells with adherent bacteria were washed and collected as above, and suspended in *Brucella* broth/10% FCS. Serial dilutions of these suspensions were spread on serum agar plates, and colonies were counted after incubation for 5 days.

Determination of IL-8 secretion

H. pylori strain P12, or its isogenic *cagT* deletion mutant, were pre-incubated with DARPins at 5 μM, cisplatin at 100 μM, or left untreated, as described above. Subsequently, AGS cells were infected with the pre-incubation mixtures at an MOI of 60, or left uninfected. Supernatants were collected after 4 h of co-incubation at 37˚C and 5% CO_2, and centrifuged to remove unbound bacteria. Aliquots of the cell supernatants were added to immunosorbent 96-well plates (Nunc MaxiSorp) that had been coated overnight with a monoclonal anti-human IL-8 capture antibody (BD Pharmingen 554716; 3 μg/ml), washed 4 times with PBS/0.05% Tween 20, and blocked for 2 h at room temperature with PBS/10% FCS. After overnight incubation at 4˚C, plates were extensively washed, then incubated with a biotinylated IL-8 detection antibody (BD Pharmingen 554718; 0.5 μg/ml) for 1 h at room temperature, and washed 4 times with PBS/0.05% Tween 20. For detection, a Vectastain ABC Kit (Biozol) was used together with a TMB substrate reagent set (BD Biosciences), according to the manufacturers's protocols. Absorbance was measured at 450 nm in a Clariostar reader, and IL-8 concentrations were calculated using a standard curve obtained with an IL-8 standard (BD Pharmingen).

Cell Adhesion and Inhibition Assays

Gastric adenocarcinoma cell line AGS (CRL 1739; American Type Culture Collection) was grown in F-12K Medium (Kaighn's Modification of Ham's F-12 Medium) supplemented with 10% fetal calf serum. Tissue culture 96-well plates (Nunc PolySorp, Dutscher, France) were coated with serial dilutions of the indicated proteins, CagL, CagI, $CagI^C$ and $CagI^N$ (0–5 μg/well in PBS) by overnight adsorption at 4˚C. The amount of adsorbed protein was determined with a BCA microprotein assay. After saturation of the wells with 1% BSA, AGS cells were collected from the culture plates by detaching with 5 mM EDTA/PBS, followed by rinsing and suspending in F-12K serum-free medium. AGS cells were seeded on ligand-coated plates at 8 x 10^4 cells/well. After 1 to 2 h, nonadherent cells were removed with a PBS wash. The extent of adhesion was determined by fixing adherent cells with 1% glutaraldehyde in PBS and then staining with 0.1% crystal violet and measuring absorbance at 570 nm as described previously

[76,77]. A blank value was subtracted that corresponded to BSA-coated wells. Each assay point was derived from triplicate measurements (three wells per assay point). Adherent cells were photographed 1 hour after seeding with an Axiovert 40 Zeiss microscope equipped with Differential Interference Contrast coupled to a Coolsnap Fx Camera (Roper Scientific, Evry, France). Cell size measurements were performed manually using Fiji 1.53c (plus) software. For AGS cell adhesion inhibition experiments with inhibitors, the coated wells were incubated with serial dilutions of DARPins in PBS for 60 min at room temperature prior to the adhesion assay as indicated in the corresponding figures. Cell adhesion data are presented as the means ± SD using Excel. For cell size measurements, a one-way Anova test was used for groups comparisons using Prism (GraphPad) software. The significance threshold was set for the t-test as $P < 0.05$. The exact sample size for replicate measurements is specified in each graph legend.

Supporting information

S1 Table. Small angle X-ray scattering data collection, processing and analysis.
(DOCX)

S1 Fig. Structural modeling of CagI dimer. Scores and sequence coverage of the AF models (left) and cartoon depiction of the AF models of CagI dimer (right) coloured according to pLDDT scores (30 to 100).
(TIF)

S2 Fig. Small Angle X-ray Scattering study of CagI, CagI^{N} and CagI^{C}. Size exclusion coupled to experimental SAXS curves of CagI (black), CagI^{N} and CagI^{C} compared to theoretical curves obtained with the corresponding model 1, 2 and 3 and domains. Theoretical curves were obtained with CagI dimers, CagI^{N} dimers and CagI^{C} monomers (chain A). Below are cartoon representations of the structures of model 1 (orange), model 2 (blue) and model 3 (magenta) fitted into the DAMMIN envelope obtained for each SAXS data. From left to right: full length CagI dimer, CagI^{N} dimer and CagI^{C} monomer.
(TIF)

S3 Fig. Sequence conservation of CagI. Clustal O alignment of CagI sequences. CagI sequence from strain 26695 (Uniprot O25273) is labelled CagI. The other CagI sequences are labelled with their uniprot code and have been chosen to illustrate the low diversity. CagL sequence corresponds to the one from strain 26695 (Uniprot O25272). Secondary structures of CagI (AF model) and CagL (PDB code 3ZCJ) are indicated above and below the alignment, respectively. The predicted signal peptide sequence (SP) is indicated by a blue box, the cysteines forming disulfide bridges are indicated by a blue dot. The arginine-glycine-aspartate [21] and D1 motifs [44] of CagL are indicated by magenta and green lines, respectively.
(TIF)

S4 Fig. Validation of binding of $_{His8}$-DARPins to CagI$_{Strep}$ by co-expression and co-purification. Coomassie blue stained SDS-PAGE of A) *E. coli* cell extracts before (-), after (+) induction of CagI$_{strep}$ expression and (E) elution fraction of His-trap column. No CagI protein could be detected in this fraction but a contaminant of around 39 kDa is visible and indicated by a *. B) co-purification of $_{His8}$-DARPins (K1-K15) with CagI-$_{Strep}$ on Ni-NTA beads showing that a band corresponding to CagI$_{strep}$ is present in the elution fraction when co-expressed with K2, K5, K9, K12, K8, K10, K11 and K15 but not with K3, K4, K6, K7, K13, K14 and K1. C) Western-blot analysis of the same samples using anti-strep antibody and stained with

NBT-BCIP. The band corresponding to $CagI_{strep}$ is indicated by an arrow.
(TIF)

S5 Fig. Sequences of DARPins identified in this study. Sequences of the DARPins, with name and clone ID indicated, aligned using ClustalO. Amino-acid differences are shaded in grey except for the single difference between K2 and K5 shaded in cyan. Repeat, N-cap and C-cap regions are indicated above the sequences.
(TIF)

S6 Fig. Affinity of DARPin for CagI and $CagI^C$. Surface Plasmon Resonance experiments using single-cycle-mode on CM5 chips coated with CagI (A) or with (B) $CagI^C$. DARPins were injected on the chips at increasing concentrations as follows. For experiments performed on full-length CagI, concentrations of DARPins were: 0.5, 2.5, 12.5, 62.5 and 312.5 nM for K9, K11, K12, K15; and 1, 3, 9, 27 and 81 nM for K10. For $CagI^C$ experiments, injections of DARPins were performed with concentrations of 0.05, 0.15, 0.45, 1.35 and 4 nM for K9, K10 and K11; 0.16, 0.8, 4, 20 and 100 nM for K12; and 2, 4, 8, 16 and 32 nM for K15. Fit curves obtained with heterogenous ligand model (CagI) or binding model 1:1 ($CagI^C$) are shown as black dashed lines.
(TIF)

S7 Fig. Determination of molar mass of CagI:DARPins complexes by size exclusion chromatography coupled to multi-angle light scattering. Each purified CagI:DARPin complex was submitted to SEC-MALS measurement represented by A280 chromatograms. Molar mass calculations are represented by dotted lines on each graph.
(TIF)

S8 Fig. Viability assessment of adherent bacteria after AGS cell infection. *H. pylori* strain P12 was treated with DARPins or cisplatin in PBS/10% FCS for 30 min at 37˚C, or left untreated, as indicated, and co-incubated with AGS cells for 2.5 h at 37˚C, 5% CO_2, Subsequently, cells were washed to remove unbound bacteria and detached by EDTA treatment, and colony-forming units (CFU) were determined by plating serial dilutions on serum agar plates. Data are indicated as normalized values (percentages in relation to CFU of untreated bacteria prior to co-incubation), and represent mean values and standard errors of the mean, resulting from three to four independent experiments.
(TIF)

S9 Fig. Quantitative measurement of cell spreading on CagL, CagI or $CagI^C$. Morphological characterization of cells adhered to 0.15 μg of CagL, CagI, $CagI^C$ 60 minutes after seeding. Cell surface area, perimeter and Feret's diameter were determined on phase contrast images using Fiji software. Each dot represents one cell (60 adhered AGS cells were measured for each condition). Each of the three parameters clearly shows that cells that have adhered to $CagI^C$ have a larger and more extended contact surface compared to cells bound to CagI or CagL. A one-way ANOVA with Tukey's post-test was used to determine significance. Means ± SD are shown with $^*p<0.1$, $^{****}p<0.0001$.
(TIF)

S10 Fig. Effects of DARPins on cell binding *cag*T4SS proteins and domains. Effect of DARPin K8, K10 or K11 on adhesion of AGS cells to CagL, CagI, $CagI^C$. Multiwell plates were coated with 4 μg of each protein. After saturation with 1% BSA, the wells were incubated with 50 μL of the indicated concentration of DARPin for 1 h at room temperature, and the cells were seeded in the presence of the inhibitor. The extent of adhesion was measured as previously described and expressed as percentage of adhesion to each protein in the absence of the

inhibitor. Each assay point was derived from triplicate measurements.
(TIF)

Acknowledgments

We acknowledge the use of the Protein Science Facility of the SFR Biosciences (UMS3444/US8). We thank Thomas Reinberg and Joana Marinho for expertly carrying out many steps in the DARPin selection and Evelyn Weiss for excellent technical assistance with CagA translocation inhibition experiments. We thank Dr. Petya Pernot from the BM29 beamline at the synchrotron ESRF for help with SAXS data collection and processing. Thanks are also due to PROXIMA 1 beamline staff for help in data collection and processing.

Author Contributions

Conceptualization: Andreas Plückthun, Sandrine Vadon-Le Goff, Patricia Rousselle, Wolfgang Fischer, Laurent Terradot.

Formal analysis: Andreas Plückthun, Sandrine Vadon-Le Goff, Patricia Rousselle, Wolfgang Fischer, Laurent Terradot.

Funding acquisition: Andreas Plückthun, Rémi Fronzes, Patricia Rousselle, Wolfgang Fischer, Laurent Terradot.

Investigation: Marine Blanc, Clara Lettl, Jérémy Guérin, Anaïs Vieille, Sven Furler, Sylvie Briand-Schumacher, Birgit Dreier, Célia Bergé, Rémi Fronzes, Patricia Rousselle, Laurent Terradot.

Writing – original draft: Marine Blanc, Laurent Terradot.

Writing – review & editing: Andreas Plückthun, Sandrine Vadon-Le Goff, Patricia Rousselle, Wolfgang Fischer, Laurent Terradot.

References

1. Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of gastric cancer attributable to Helicobacter pylori. Int J Cancer. 2015; 136(2):487–90. Epub 2014/06/04. https://doi.org/10.1002/ijc.28999 PMID: 24889903.
2. Park JY, Forman D, Waskito LA, Yamaoka Y, Crabtree JE. Epidemiology of Helicobacter pylori and CagA-Positive Infections and Global Variations in Gastric Cancer. Toxins (Basel). 2018; 10(4). Epub 2018/04/20. https://doi.org/10.3390/toxins10040163 PMID: 29671784; PubMed Central PMCID: PMC5923329.
3. Knorr J, Ricci V, Hatakeyama M, Backert S. Classification of Helicobacter pylori Virulence Factors: Is CagA a Toxin or Not? Trends Microbiol. 2019; 27(9):731–8. Epub 2019/05/28. https://doi.org/10.1016/j.tim.2019.04.010 PMID: 31130493.
4. Cover TL, Lacy DB, Ohi MD. The Helicobacter pylori Cag Type IV Secretion System. Trends Microbiol. 2020; 28(8):682–95. Epub 2020/05/27. https://doi.org/10.1016/j.tim.2020.02.004 PMID: 32451226; PubMed Central PMCID: PMC7363556.
5. Waksman G. From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. EMBO Rep. 2019; 20(2). Epub 2019/01/04. https://doi.org/10.15252/embr.201847012 PMID: 30602585; PubMed Central PMCID: PMC6362355.
6. Costa TRD, Harb L, Khara P, Zeng L, Hu B, Christie PJ. Type IV secretion systems: Advances in structure, function, and activation. Mol Microbiol. 2021; 115(3):436–52. Epub 2020/12/17. https://doi.org/10.1111/mmi.14670 PMID: 33326642; PubMed Central PMCID: PMC8026593.
7. Low HH, Gubellini F, Rivera-Calzada A, Braun N, Connery S, Dujeancourt A, et al. Structure of a type IV secretion system. Nature. 2014; 508(7497):550–3. Epub 2014/03/29. https://doi.org/10.1038/nature13081 [pii]. PMID: 24670658; PubMed Central PMCID: PMC3998870.

8. Costa TR, Ilangovan A, Ukleja M, Redzej A, Santini JM, Smith TK, et al. Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. Cell. 2016; 166 (6):1436–44 e10. https://doi.org/10.1016/j.cell.2016.08.025 PMID: 27610568; PubMed Central PMCID: PMC5018250.

9. Redzej A, Ukleja M, Connery S, Trokter M, Felisberto-Rodrigues C, Cryar A, et al. Structure of a VirD4 coupling protein bound to a VirB type IV secretion machinery. EMBO J. 2017. https://doi.org/10.15252/embj.201796629 PMID: 28923826.

10. Sheedlo MJ, Ohi MD, Lacy DB, Cover TL. Molecular architecture of bacterial type IV secretion systems. PLoS Pathog. 2022; 18(8):e1010720. Epub 2022/08/12. https://doi.org/10.1371/journal.ppat.1010720 PMID: 35951533; PubMed Central PMCID: PMC9371333.

11. Sheedlo MJ, Chung JM, Sawhney N, Durie CL, Cover TL, Ohi MD, et al. Cryo-EM reveals species-specific components within the Helicobacter pylori Cag type IV secretion system core complex. Elife. 2020; 9. Epub 2020/09/03. https://doi.org/10.7554/eLife.59495 PMID: 32876048; PubMed Central PMCID: PMC7511236.

12. Tanaka J, Suzuki T, Mimuro H, Sasakawa C. Structural definition on the surface of Helicobacter pylori type IV secretion apparatus. Cell Microbiol. 2003; 5(6):395–404. https://doi.org/10.1046/j.1462-5822.2003.00286.x PMID: 12780777.

13. Rohde M, Puls J, Buhrdorf R, Fischer W, Haas R. A novel sheathed surface organelle of the Helicobacter pylori cag type IV secretion system. Mol Microbiol. 2003; 49(1):219–34. https://doi.org/10.1046/j.1365-2958.2003.03549.x PMID: 12823823.

14. Chang YW, Shaffer CL, Rettberg LA, Ghosal D, Jensen GJ. In Vivo Structures of the Helicobacter pylori cag Type IV Secretion System. Cell Rep. 2018; 23(3):673–81. https://doi.org/10.1016/j.celrep.2018.03.085 PMID: 29669273.

15. Johnson EM, Gaddy JA, Voss BJ, Hennig EE, Cover TL. Genes required for assembly of pili associated with the Helicobacter pylori cag type IV secretion system. Infect Immun. 2014; 82(8):3457–70. Epub 2014/06/04. https://doi.org/10.1128/IAI.01640-14 [pii]. PMID: 24891108; PubMed Central PMCID: PMC4136203.

16. Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, Hennig EE, et al. Helicobacter pylori exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface. PLoS Pathog. 2011; 7(9):e1002237. Epub 2011/09/13. https://doi.org/10.1371/journal.ppat.1002237 [pii]. PMID: 21909278; PubMed Central PMCID: PMC3164655.

17. Fischer W, Puls J, Buhrdorf R, Gebert B, Odenbreit S, Haas R. Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol. 2001; 42(5):1337–48. Epub 2002/03/12. https://doi.org/10.1046/j.1365-2958.2001.02714.x PMID: 11886563.

18. Berge C, Terradot L. Structural Insights into Helicobacter pylori Cag Protein Interactions with Host Cell Factors. Curr Top Microbiol Immunol. 2017; 400:129–47. https://doi.org/10.1007/978-3-319-50520-6_6 PMID: 28124152.

19. Barden S, Lange S, Tegtmeyer N, Conradi J, Sewald N, Backert S, et al. A helical RGD motif promoting cell adhesion: crystal structures of the Helicobacter pylori type IV secretion system pilus protein CagL. Structure. 2013; 21(11):1931–41. https://doi.org/10.1016/j.str.2013.08.018 PMID: 24076404.

20. Tegtmeyer N, Hartig R, Delahay RM, Rohde M, Brandt S, Conradi J, et al. A small fibronectin-mimicking protein from bacteria induces cell spreading and focal adhesion formation. J Biol Chem. 2010; 285 (30):23515–26. Epub 2010/05/29. https://doi.org/10.1074/jbc.M109.096214 [pii] PMID: 20507990; PubMed Central PMCID: PMC2906342.

21. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, et al. Helicobacter exploits integrin for type IV secretion and kinase activation. Nature. 2007; 449(7164):862–6. https://doi.org/10.1038/nature06187 PMID: 17943123.

22. Buss M, Tegtmeyer N, Schnieder J, Dong X, Li J, Springer TA, et al. Specific high affinity interaction of Helicobacter pylori CagL with integrin alphaV beta6 promotes type IV secretion of CagA into human cells. FEBS J. 2019; 286(20):3980–97. Epub 2019/06/15. https://doi.org/10.1111/febs.14962 PMID: 31197920.

23. Choi JM, Choi YH, Sudhanva MS, Devakumar S, Lee KH, Cha JH, et al. Crystal structure of CagL from Helicobacter pylori K74 strain. Biochem Biophys Res Commun. 2015; 460(4):964–70. https://doi.org/10.1016/j.bbrc.2015.03.135 PMID: 25839651.

24. Bonsor DA, Pham KT, Beadenkopf R, Diederichs K, Haas R, Beckett D, et al. Integrin engagement by the helical RGD motif of the Helicobacter pylori CagL protein is regulated by pH-induced displacement of a neighboring helix. J Biol Chem. 2015; 290(20):12929–40. https://doi.org/10.1074/jbc.M115.641829 PMID: 25837254; PubMed Central PMCID: PMC4432307.

25. Barden S, Schomburg B, Conradi J, Backert S, Sewald N, Niemann HH. Structure of a three-dimensional domain-swapped dimer of the Helicobacter pylori type IV secretion system pilus protein CagL. Acta Crystallogr D Biol Crystallogr. 2014; 70(Pt 5):1391–400. https://doi.org/10.1107/S1399004714003150 PMID: 24816107.

26. Bonig T, Olbermann P, Bats SH, Fischer W, Josenhans C. Systematic site-directed mutagenesis of the Helicobacter pylori CagL protein of the Cag type IV secretion system identifies novel functional domains. Sci Rep. 2016; 6:38101. https://doi.org/10.1038/srep38101 PMID: 27922023; PubMed Central PMCID: PMC5138618.

27. Backert S, Fronzes R, Waksman G. VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? Trends Microbiol. 2008; 16(9):409–13. https://doi.org/10.1016/j.tim.2008.07.001 PMID: 18706815.

28. Koelblen T, Berge C, Cherrier MV, Brillet K, Jimenez-Soto L, Ballut L, et al. Molecular dissection of protein-protein interactions between integrin alpha5beta1 and the Helicobacter pylori Cag Type IV secretion system. FEBS J. 2017. https://doi.org/10.1111/febs.14299 PMID: 29055076.

29. Jimenez-Soto LF, Kutter S, Sewald X, Ertl C, Weiss E, Kapp U, et al. Helicobacter pylori type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner. PLoS Pathog. 2009; 5(12): e1000684. Epub 2009/12/10. https://doi.org/10.1371/journal.ppat.1000684 PMID: 19997503; PubMed Central PMCID: PMC2779590.

30. Kumar N, Shariq M, Kumari R, Tyagi RK, Mukhopadhyay G. Cag type IV secretion system: CagI independent bacterial surface localization of CagA. PLoS One. 2013; 8(9):e74620. https://doi.org/10.1371/journal.pone.0074620 PMID: 24040297; PubMed Central PMCID: PMC3769253.

31. Pham KT, Weiss E, Jimenez Soto LF, Breithaupt U, Haas R, Fischer W. CagI is an essential component of the Helicobacter pylori Cag type IV secretion system and forms a complex with CagL. PLoS One. 2012; 7(4):e35341. https://doi.org/10.1371/journal.pone.0035341 PMID: 22493745; PubMed Central PMCID: PMC3320882.

32. Kumar N, Shariq M, Kumar A, Kumari R, Subbarao N, Tyagi RK, et al. Analyzing the role of CagV, a VirB8 homolog of the type IV secretion system of Helicobacter pylori. FEBS Open Bio. 2017; 7(7):915–33. Epub 2017/07/07. https://doi.org/10.1002/2211-5463.12225 PMID: 28680806; PubMed Central PMCID: PMC5494299.

33. Pluckthun A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. Annu Rev Pharmacol Toxicol. 2015; 55:489–511. Epub 2015/01/07. https://doi.org/10.1146/annurev-pharmtox-010611-134654 PMID: 25562645.

34. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873):583–9. Epub 2021/07/16. https://doi.org/10.1038/s41586-021-03819-2 PMID: 34265844; PubMed Central PMCID: PMC8371605.

35. Piiadov V, Ares de Araujo E, Oliveira Neto M, Craievich AF, Polikarpov I. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. Protein Sci. 2019; 28(2):454–63. Epub 2018/10/30. https://doi.org/10.1002/pro.3528 PMID: 30371978; PubMed Central PMCID: PMC6319763.

36. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372(3):774–97. Epub 2007/08/08. https://doi.org/10.1016/j.jmb.2007.05.022 PMID: 17681537.

37. Schindele F, Weiss E, Haas R, Fischer W. Quantitative analysis of CagA type IV secretion by Helicobacter pylori reveals substrate recognition and translocation requirements. Mol Microbiol. 2016; 100 (1):188–203. https://doi.org/10.1111/mmi.13309 PMID: 26713727.

38. Lettl C, Schindele F, Testolin G, Bar A, Rehm T, Bronstrup M, et al. Inhibition of Type IV Secretion Activity and Growth of Helicobacter pylori by Cisplatin and Other Platinum Complexes. Front Cell Infect Microbiol. 2020; 10:602958. Epub 2021/01/05. https://doi.org/10.3389/fcimb.2020.602958 PMID: 33392108; PubMed Central PMCID: PMC7775389.

39. Lettl C, Haas R, Fischer W. Kinetics of CagA type IV secretion by Helicobacter pylori and the requirement for substrate unfolding. Mol Microbiol. 2021; 116(3):794–807. Epub 2021/06/15. https://doi.org/10.1111/mmi.14772 PMID: 34121254.

40. Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, et al. Age of the association between Helicobacter pylori and man. PLoS Pathog. 2012; 8(5):e1002693. Epub 2012/05/17. https://doi.org/10.1371/journal.ppat.1002693 PMID: 22589724; PubMed Central PMCID: PMC3349757.

41. Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, et al. An African origin for the intimate association between humans and Helicobacter pylori. Nature. 2007; 445(7130):915–8. https://doi.org/10.1038/nature05562 PMID: 17287725.

42. Ailloud F, Didelot X, Woltemate S, Pfaffinger G, Overmann J, Bader RC, et al. Within-host evolution of Helicobacter pylori shaped by niche-specific adaptation, intragastric migrations and selective sweeps.

Nat Commun. 2019; 10(1):2273. Epub 2019/05/24. https://doi.org/10.1038/s41467-019-10050-1 PMID: 31118420; PubMed Central PMCID: PMC6531487.

43. Tegtmeyer N, Neddermann M, Lind J, Pachathundikandi SK, Sharafutdinov I, Gutierrez-Escobar AJ, et al. Toll-like Receptor 5 Activation by the CagY Repeat Domains of Helicobacter pylori. Cell Rep. 2020; 32(11):108159. Epub 2020/09/17. https://doi.org/10.1016/j.celrep.2020.108159 PMID: 32937132.

44. Pachathundikandi SK, Tegtmeyer N, Arnold IC, Lind J, Neddermann M, Falkeis-Veits C, et al. T4SS-dependent TLR5 activation by Helicobacter pylori infection. Nat Commun. 2019; 10(1):5717. Epub 2019/12/18. https://doi.org/10.1038/s41467-019-13506-6 PMID: 31844047; PubMed Central PMCID: PMC6915727.

45. Hayashi T, Senda M, Morohashi H, Higashi H, Horio M, Kashiba Y, et al. Tertiary structure-function analysis reveals the pathogenic signaling potentiation mechanism of Helicobacter pylori oncogenic effector CagA. Cell Host Microbe. 2012; 12(1):20–33. Epub 2012/07/24. https://doi.org/10.1016/j.chom.2012.05.010 PMID: 22817985.

46. Kaplan-Turkoz B, Jimenez-Soto LF, Dian C, Ertl C, Remaut H, Louche A, et al. Structural insights into Helicobacter pylori oncoprotein CagA interaction with beta1 integrin. Proc Natl Acad Sci U S A. 2012; 109(36):14640–5. Epub 2012/08/22. https://doi.org/10.1073/pnas.1206098109 PMID: 22908298; PubMed Central PMCID: PMC3437852.

47. Chung JM, Sheedlo MJ, Campbell AM, Sawhney N, Frick-Cheng AE, Lacy DB, et al. Structure of the Helicobacter pylori Cag type IV secretion system. Elife. 2019; 8. https://doi.org/10.7554/eLife.47644 PMID: 31210639; PubMed Central PMCID: PMC6620104.

48. Aly KA, Baron C. The VirB5 protein localizes to the T-pilus tips in Agrobacterium tumefaciens. Microbiology (Reading). 2007; 153(Pt 11):3766–75. Epub 2007/11/03. https://doi.org/10.1099/mic.0.2007/010462-0 PMID: 17975085.

49. Yeo HJ, Yuan Q, Beck MR, Baron C, Waksman G. Structural and functional characterization of the VirB5 protein from the type IV secretion system encoded by the conjugative plasmid pKM101. Proc Natl Acad Sci U S A. 2003; 100(26):15947–52. https://doi.org/10.1073/pnas.2535211100 PMID: 14673074; PubMed Central PMCID: PMC307673.

50. Gonzalez-Rivera C, Khara P, Awad D, Patel R, Li YG, Bogisch M, et al. 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. 2019; 111(1):96–117. Epub 2018/09/29. https://doi.org/10.1111/mmi.14141 PMID: 30264928; PubMed Central PMCID: PMC6351158.

51. Zhao Q, Busch B, Jimenez-Soto LF, Ishikawa-Ankerhold H, Massberg S, Terradot L, et al. Integrin but not CEACAM receptors are dispensable for Helicobacter pylori CagA translocation. PLoS Pathog. 2018; 14(10):e1007359. https://doi.org/10.1371/journal.ppat.1007359 PMID: 30365569; PubMed Central PMCID: PMC6231679.

52. Tegtmeyer N, Wessler S, Necchi V, Rohde M, Harrer A, Rau TT, et al. Helicobacter pylori Employs a Unique Basolateral Type IV Secretion Mechanism for CagA Delivery. Cell Host Microbe. 2017; 22(4):552–60 e5. Epub 2017/10/13. https://doi.org/10.1016/j.chom.2017.09.005 PMID: 29024645.

53. Steadman D, Lo A, Waksman G, Remaut H. Bacterial surface appendages as targets for novel antibacterial therapeutics. Future Microbiol. 2014; 9:887–900. Epub 2014/08/27. https://doi.org/10.2217/fmb.14.46 PMID: 25156378.

54. Hotinger JA, Pendergrass HA, May AE. Molecular Targets and Strategies for Inhibition of the Bacterial Type III Secretion System (T3SS); Inhibitors Directly Binding to T3SS Components. Biomolecules. 2021; 11(2). Epub 2021/03/07. https://doi.org/10.3390/biom11020316 PMID: 33669653; PubMed Central PMCID: PMC7922566.

55. Boudaher E, Shaffer CL. Inhibiting bacterial secretion systems in the fight against antibiotic resistance. Medchemcomm. 2019; 10(5):682–92. Epub 2019/11/20. https://doi.org/10.1039/c9md00076c PMID: 31741728; PubMed Central PMCID: PMC6677025.

56. Debraekeleer A, Remaut H. Future perspective for potential Helicobacter pylori eradication therapies. Future Microbiol. 2018; 13:671–87. Epub 2018/05/26. https://doi.org/10.2217/fmb-2017-0115 PMID: 29798689.

57. Sayer JR, Wallden K, Koss H, Allan H, Daviter T, Gane PJ, et al. Design, synthesis, and evaluation of peptide-imidazo[1,2-a]pyrazine bioconjugates as potential bivalent inhibitors of the VirB11 ATPase HP0525. J Pept Sci. 2021; 27(10):e3353. Epub 2021/06/19. https://doi.org/10.1002/psc.3353 PMID: 34142414.

58. Arya T, Oudouhou F, Casu B, Bessette B, Sygusch J, Baron C. Fragment-based screening identifies inhibitors of ATPase activity and of hexamer formation of Cagalpha from the Helicobacter pylori type IV secretion system. Sci Rep. 2019; 9(1):6474. Epub 2019/04/26. https://doi.org/10.1038/s41598-019-42876-6 PMID: 31019200; PubMed Central PMCID: PMC6482174.

59. Shaffer CL, Good JA, Kumar S, Krishnan KS, Gaddy JA, Loh JT, et al. Peptidomimetic Small Molecules Disrupt Type IV Secretion System Activity in Diverse Bacterial Pathogens. mBio. 2016; 7(2):e00221–16. Epub 2016/04/28. https://doi.org/10.1128/mBio.00221-16 PMID: 27118587; PubMed Central PMCID: PMC4850256.

60. Faass L, Hauke M, Stein SC, Josenhans C. Innate immune activation and modulatory factors of Helicobacter pylori towards phagocytic and nonphagocytic cells. Curr Opin Immunol. 2023; 82:102301. Epub 2023/03/19. https://doi.org/10.1016/j.coi.2023.102301 PMID: 36933362.

61. Evans R O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2022:2021.10.04.463034. https://doi.org/10.1101/2021.10.04.463034

62. Fairhead M, Howarth M. Site-specific biotinylation of purified proteins using BirA. Methods Mol Biol. 2015; 1266:171–84. Epub 2015/01/07. https://doi.org/10.1007/978-1-4939-2272-7_12 PMID: 25560075; PubMed Central PMCID: PMC4304673.

63. Binz HK, Stumpp MT, Forrer P, Amstutz P, Pluckthun A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol. 2003; 332(2):489–503. Epub 2003/09/02. https://doi.org/10.1016/s0022-2836(03)00896-9 PMID: 12948497.

64. Brauchle M, Hansen S, Caussinus E, Lenard A, Ochoa-Espinosa A, Scholz O, et al. Protein interference applications in cellular and developmental biology using DARPins that recognize GFP and mCherry. Biol Open. 2014; 3(12):1252–61. Epub 2014/11/25. https://doi.org/10.1242/bio.201410041 PMID: 25416061; PubMed Central PMCID: PMC4265764.

65. Kramer MA, Wetzel SK, Pluckthun A, Mittl PR, Grutter MG. Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module. J Mol Biol. 2010; 404 (3):381–91. Epub 2010/09/21. https://doi.org/10.1016/j.jmb.2010.09.023 PMID: 20851127.

66. Schilling J, Schoppe J, Pluckthun A. From DARPins to LoopDARPins: novel LoopDARPin design allows the selection of low picomolar binders in a single round of ribosome display. J Mol Biol. 2014; 426 (3):691–721. Epub 2014/02/12. https://doi.org/10.1016/j.jmb.2013.10.026 PMID: 24513107.

67. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. Journal of Applied Crystallography. 1993; 26:795–800.

68. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013; 69(Pt 7):1204–14. Epub 2013/06/26. https://doi.org/10.1107/S0907444913000061 PMID: 23793146; PubMed Central PMCID: PMC3689523.

69. Collaborative Computational Project-4. The CCP4 suite: Programs for protein crystallography. Acta Crystallogr. 1994;D50:760–3.

70. Hough MA, Wilson KS. From crystal to structure with CCP4. Acta Crystallogr D Struct Biol. 2018; 74(Pt 2):67. Epub 2018/03/14. https://doi.org/10.1107/S2059798317017557 PMID: 29533232; PubMed Central PMCID: PMC5947770.

71. Vigano MA, Bieli D, Schaefer JV, Jakob RP, Matsuda S, Maier T, et al. DARPins recognizing mTFP1 as novel reagents for in vitro and in vivo protein manipulations. Biol Open. 2018; 7(11). Epub 2018/09/22. https://doi.org/10.1242/bio.036749 PMID: 30237292; PubMed Central PMCID: PMC6262872.

72. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40(Pt 4):658–74. Epub 2007/08/01. https://doi.org/10.1107/S0021889807021206 PMID: 19461840; PubMed Central PMCID: PMC2483472.

73. Petoukhov MV, Svergun DI. Applications of small-angle X-ray scattering to biomacromolecular solutions. Int J Biochem Cell Biol. 2013; 45(2):429–37. Epub 2012/11/13. https://doi.org/10.1016/j.biocel.2012.10.017 PMID: 23142499.

74. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J. 2013; 105(4):962–74. Epub 2013/08/27. https://doi.org/10.1016/j.bpj.2013.07.020 PMID: 23972848; PubMed Central PMCID: PMC3752106.

75. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. Biophys J. 2001; 80(6):2946–53. Epub 2001/05/24. https://doi.org/10.1016/S0006-3495(01)76260-1 PMID: 11371467; PubMed Central PMCID: PMC1301478.

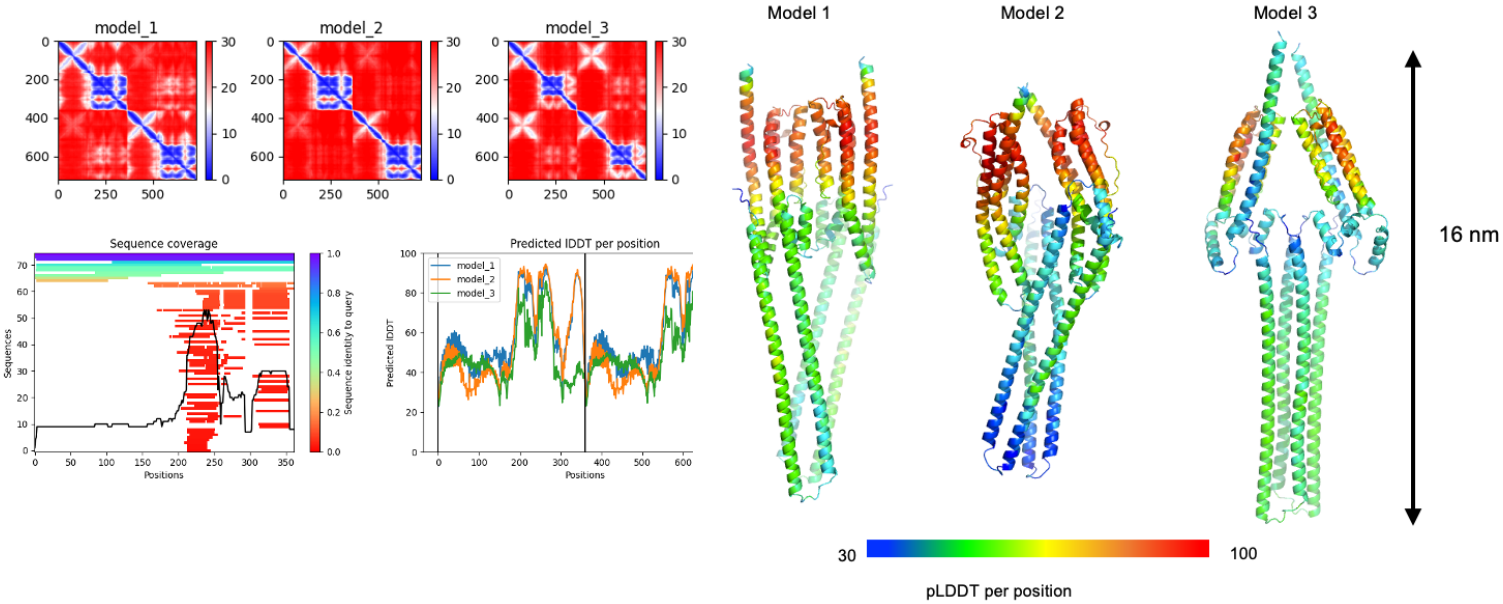
76. Carulli S, Beck K, Dayan G, Boulesteix S, Lortat-Jacob H, Rousselle P. Cell surface proteoglycans syndecan-1 and -4 bind overlapping but distinct sites in laminin alpha3 LG45 protein domain. J Biol Chem. 2012; 287(15):12204–16. Epub 2012/02/22. https://doi.org/10.1074/jbc.M111.300061 PMID: 22351752; PubMed Central PMCID: PMC3320972.

77. Sulka B, Lortat-Jacob H, Terreux R, Letourneur F, Rousselle P. Tyrosine dephosphorylation of the syndecan-1 PDZ binding domain regulates syntenin-1 recruitment. J Biol Chem. 2009; 284(16):10659–71. Epub 2009/02/21. https://doi.org/10.1074/jbc.M807643200 PMID: 19228696; PubMed Central PMCID: PMC2667753.

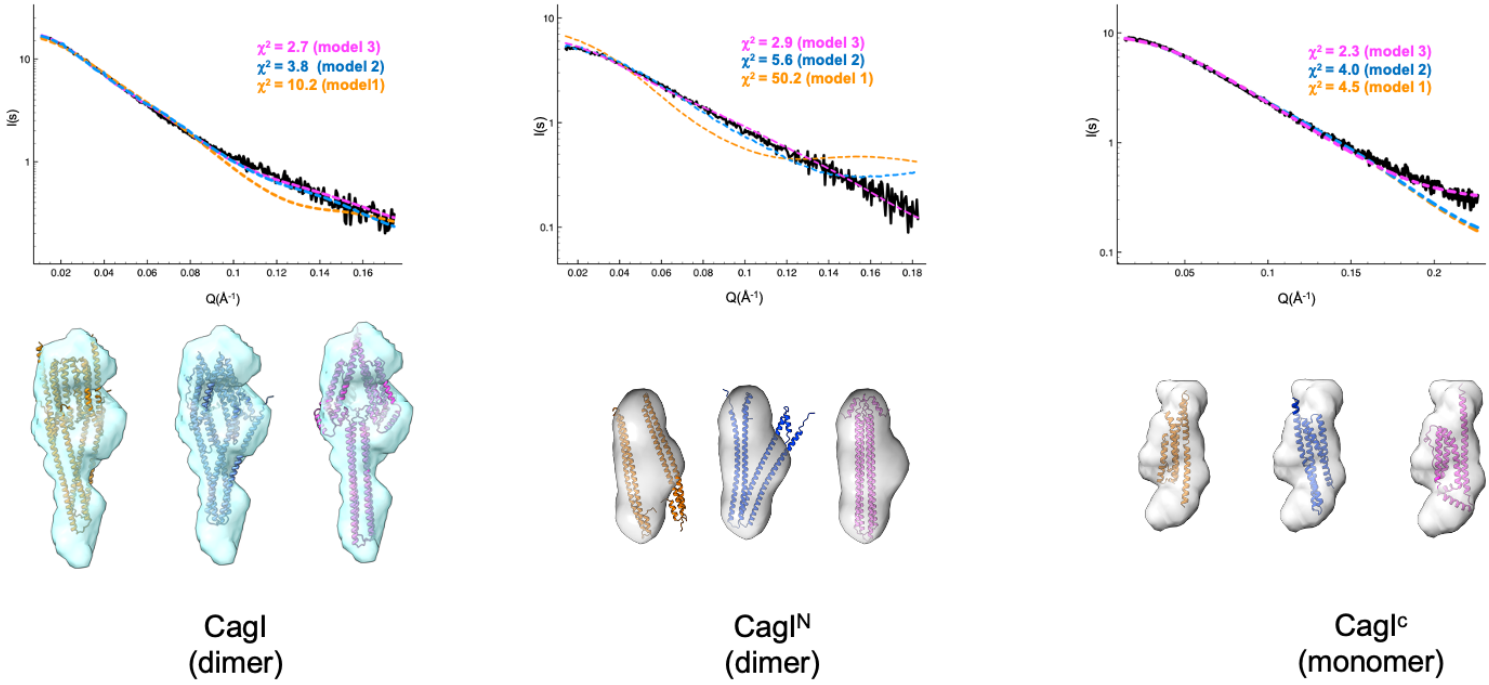
S1 Table. Small angle X-ray scattering data collection, processing and analysis

	CagI	$CagI^N$	$CagI^C$
Data-collection parameters:			
Q range ($Å^{-1}$)	0.01-0.17	0.01-0.18	0.01-0.23
Structural parameters:			
I(0) ($Å^{-1}$) (from P(r))	18.2	5.7	9.4
Rg (Å) (from P(r))	4.66 ± 0.03	2.9 ± 0.1	2.4 ± 0.02
I(0) ($Å^{-1}$) (from Guinier)	18.2	5.7	9.4
Rg (Å) (from Guinier)	4.65 ± 0.07	2.9 ± 0.1	2.3 ± 0.02
Dmax (Å)	162	116	88
Porod volume estimation. V_0 ($Å^3$)	155.3	54.4	27.8
Molecular-mass determination:			
Molecular mass Mr (kDa) (SaxsMOW)	99.1	32	20.3
Molecular mass Mr (kDa) (Porod)	91.3	32	16.3
Calculated mass Mr (kDa) (sequence)	78.2	35.6	18.1

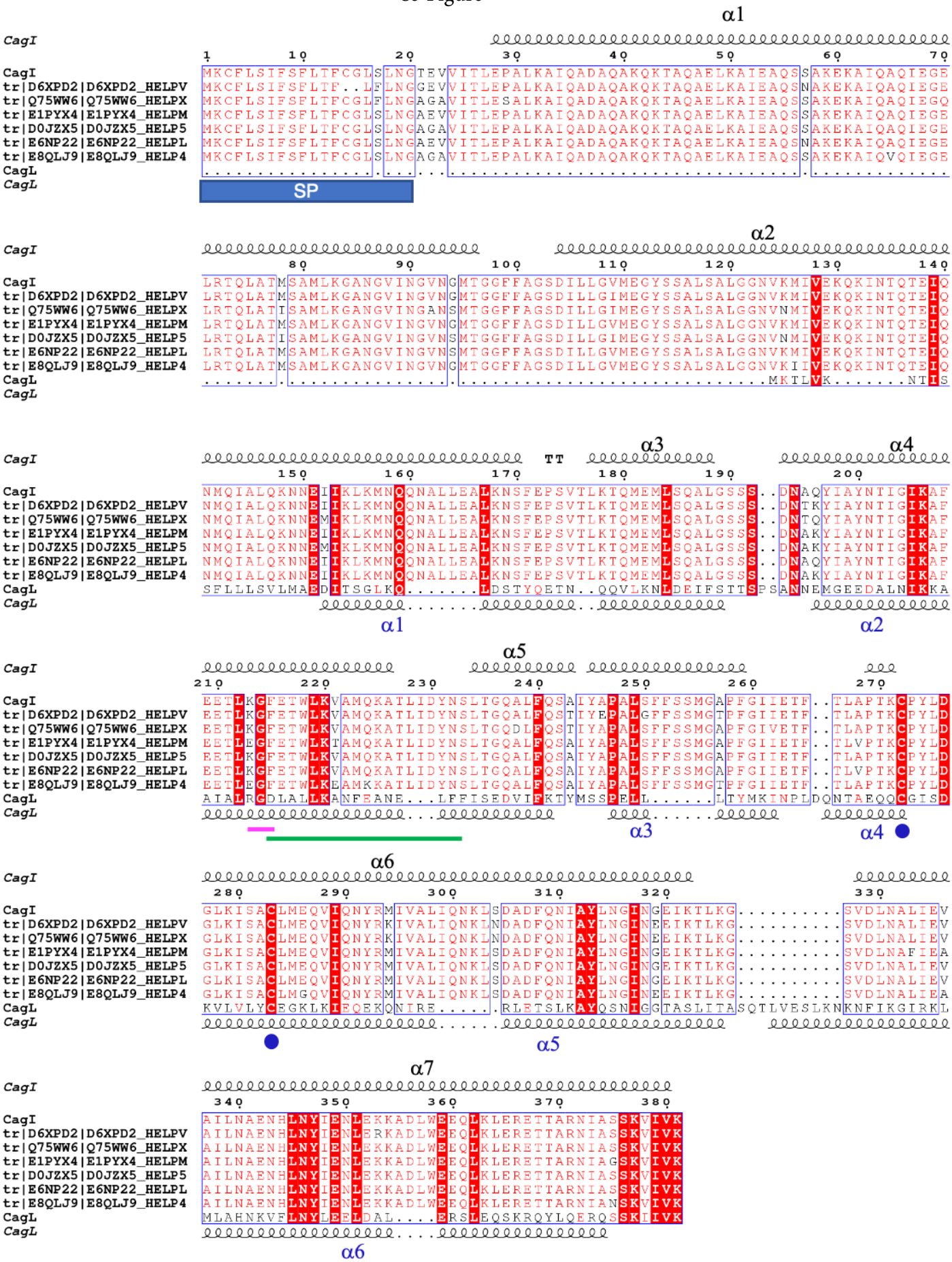
S1 Figure


model_1
model_2
model_3
Sequence coverage
Sequences
Positions
Sequence identity to query
Predicted lDDT per position
Predicted lDDT
Positions
model_1
model_2
model_3
Model 1
Model 2
Model 3
16 nm
30
100
pLDDT per position

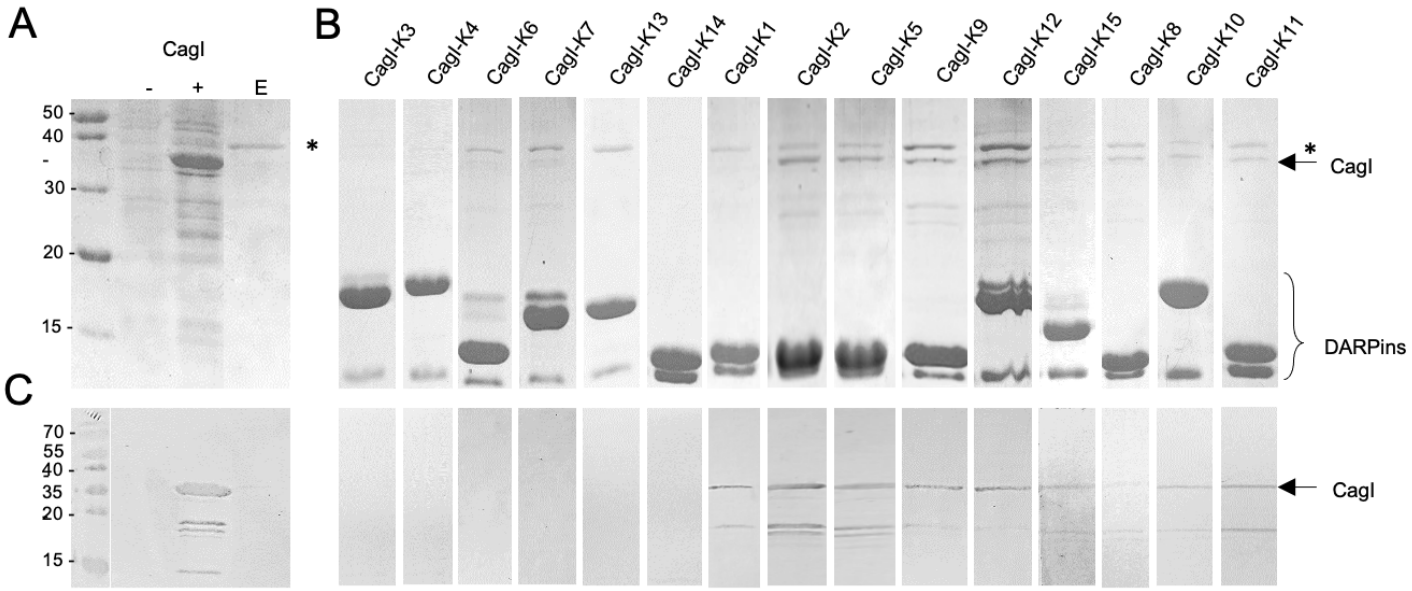
S2 Figure


χ² = 2.7 (model 3)
χ² = 3.8 (model 2)
χ² = 10.2 (model1)
I(s)
Q(Å⁻¹)
χ² = 2.9 (model 3)
χ² = 5.6 (model 2)
χ² = 50.2 (model 1)
I(s)
Q(Å⁻¹)
χ² = 2.3 (model 3)
χ² = 4.0 (model 2)
χ² = 4.5 (model 1)
I(s)
Q(Å⁻¹)
CagI
(dimer)
CagI^N
(dimer)
CagI^C
(monomer)

S3 Figure


α1
CagI
1 10 20 30 40 50 60 70
CagI
tr|D6XPD2|D6XPD2_HELPV
tr|Q75WW6|Q75WW6_HELPX
tr|E1PYX4|E1PYX4_HELPM
tr|D0JZX5|D0JZX5_HELP5
tr|E6NP22|E6NP22_HELPL
tr|E8QLJ9|E8QLJ9_HELP4
CagL
CagL
SP
α2
80 90 100 110 120 130 140
α3
α4
150 160 170 180 190 200
α1
α2
α5
210 220 230 240 250 260 270
α3
α4
α6
280 290 300 310 320 330
α5
α7
340 350 360 370 380
α6

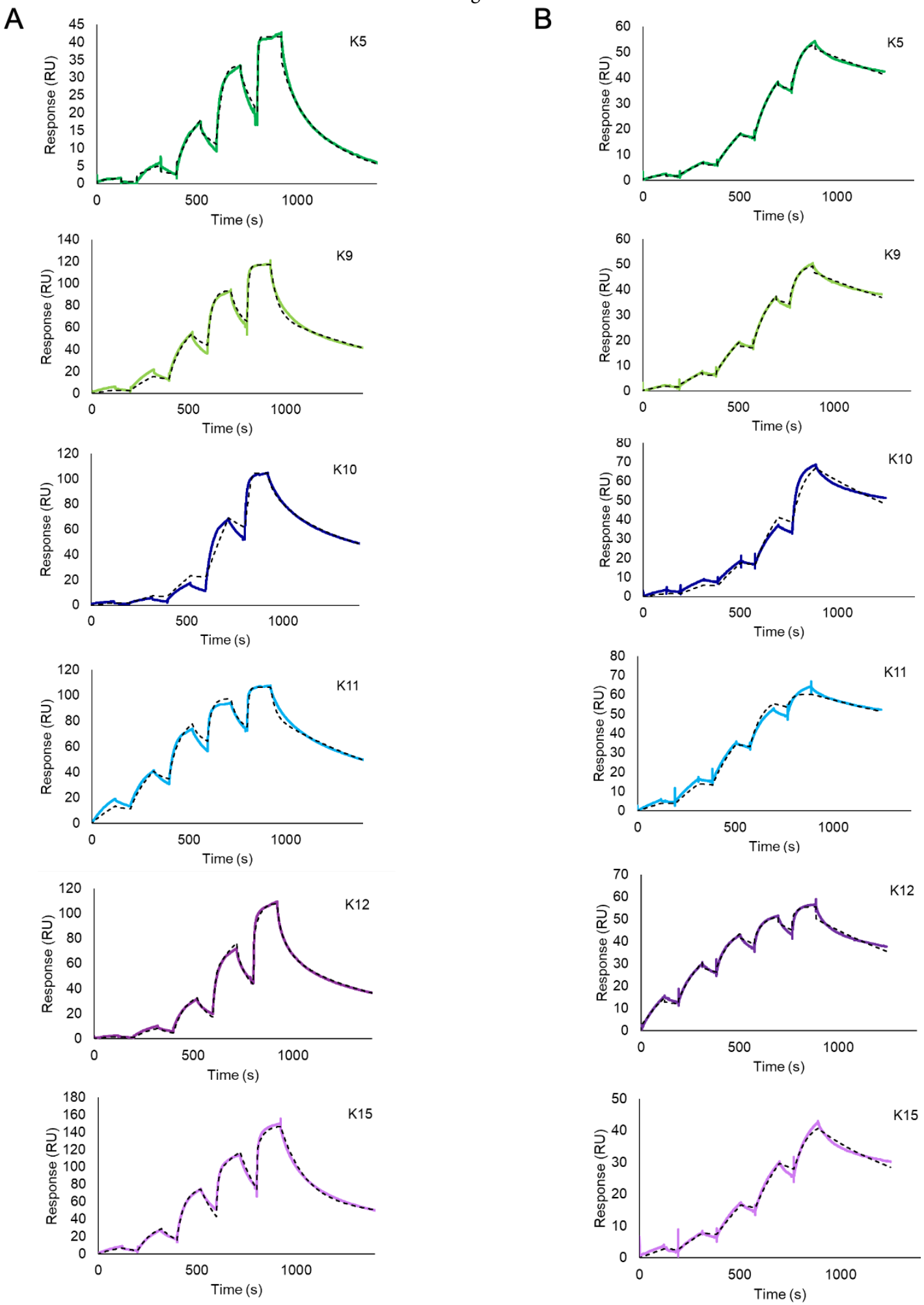
S4 Figure


A
CagI
-
+
E
50
40
30
20
15
*
B
CagI-K3
CagI-K4
CagI-K6
CagI-K7
CagI-K13
CagI-K14
CagI-K1
CagI-K2
CagI-K5
CagI-K9
CagI-K12
CagI-K15
CagI-K8
CagI-K10
CagI-K11
*
CagI
DARPins
C
70
55
40
35
20
15
CagI

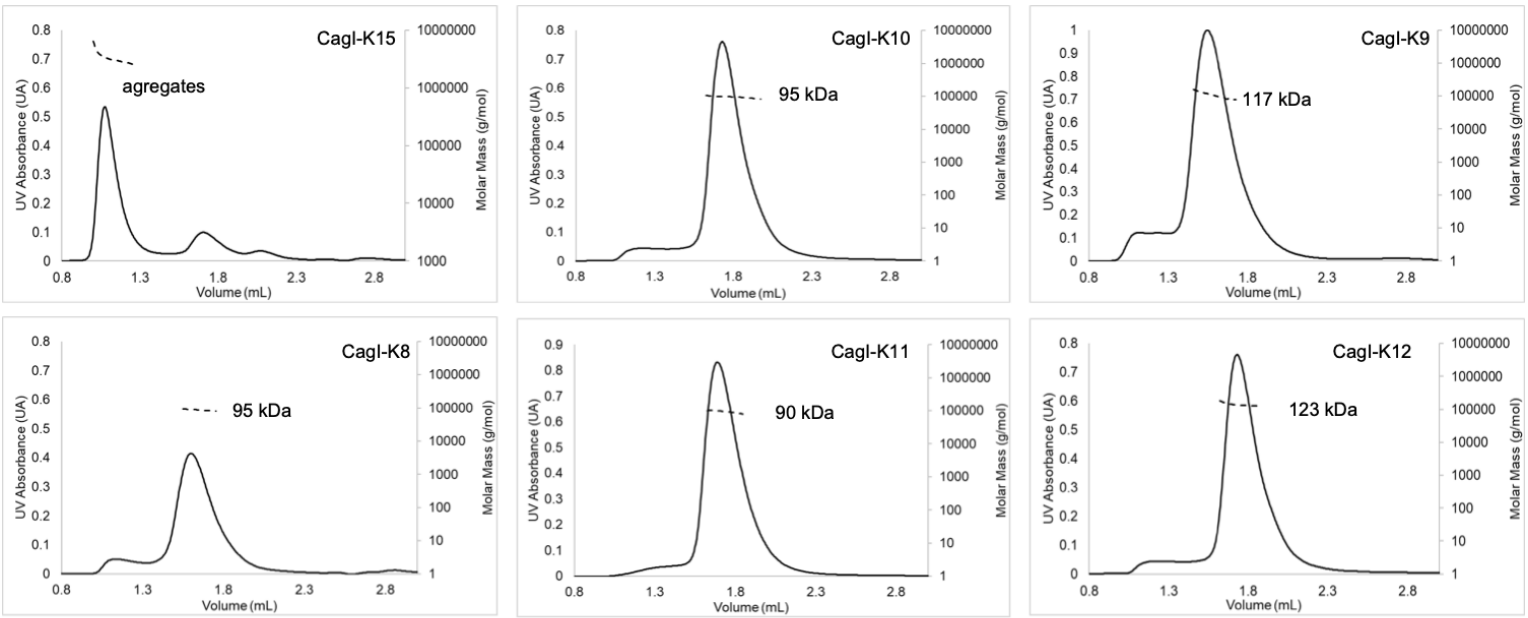
S5 Figure


DARPin
Clone ID
N-cap
1st repeat
K2 008-801-2176-F2 DLGKKLLEAA LIGQDDEVRI LMANGADVNA MDNFGHTPLH LAAMMGHLEI VEVLLKTGAD VNA
K5 008-801-2177-B8 DLGKKLLEAA LIGQDDEVRI LMANGADVNA MDNFGHTPLH LAAMMGHLEI VEVLLKTGAD VNA
K8 008-801-2177-G11 DLGKKLLEAA TIGQHDEVRI LMANGADVNA WDFLGQTPLH LAANMGHLEI VEVLLKAGAD VNA
K9 008-801-2178-B4 DLGKKLLEAA RAGQDDEVRI LMANGADVNA TDMAGWTPLH LAAIEGHLEI VEVLLKTGAD VNA
K10 008-801-2178-C11 DLGKKLLEAA VQGQDDEVRI LMANGADVNA QDQAGHTPLH LAAVRGHLEI VEVLLKTGAD VNA
K11 008-801-2178-E11 DLGKKLLEAA LYGQDDEVRI LMANGADVNA QDQAGHTPLH LAALVGHLEI VEVLLKTGAD VNA
K12 008-801-2179-E4 DLGKKLLEAA HAGQDDEVRI LMANGADVNA SDIWGFTPLH LAALWGHLEI VEVLLKTGAD VNA
K15 008-801-2179-E12 DLGKKLLEAA SWGQDDEVRI LMANGADVE- ----GYTPLH LAASWGHLEI VEVLLKTGAD VNA
2nd repeat
3rd repeat
K2 008-801-2176-F2 FDLTGFT PLHLAAYAGH LEIVEVLLKH GADVNA---- ---------- ---------- ----------
K5 008-801-2177-B8 FDLTGFT PLHLAAYAGH LEIVEVLLKH GADVNA---- ---------- ---------- ----------
K8 008-801-2177-G11 EDNHGFT PLHLAAYWGH LEIVEVLLKH GADVNA---- ---------- ---------- ----------
K9 008-801-2178-B4 EDAFGLT PLHLAAWFGH LEIVEVLLKH GADVNA---- ---------- ---------- ----------
K10 008-801-2178-C11 TDDAGWT PLHLAAMHGH LEIVEVLLKA GADVNATDVW GHTPLHLVAV REGHLEIVEV LLKHGA DVNA
K11 008-801-2178-E11 ADDFGFT PLHLAAFYGH LEIVEVLLKH GADVNA---- ---------- ---------- ----------
K12 008-801-2179-E4 NDSFGET PLHLAASFGH LEIVEVLLKA GADVNAIDYF GWTPLHLAAV -NGHLEIVEV LLKTGA DVNA
K15 008-801-2179-E12 HDMNGFT PLHLAAFYGH LEIVEVLLNA GADVNAQDYQ GNTPLHLAAM -FGHLEIVEV LLKHGA DVNA
C-cap
K2 008-801-2176-F2 QDQDGATPFD LAAWFGNEDI AEVLQKAAKL
K5 008-801-2177-B8 QDQDGATPFD LAAWLGNEDI AEVLQKAAKL
K8 008-801-2177-G11 QDATGVTPFD LAAYYGNEDI AEVLQKAAKL
K9 008-801-2178-B4 QDKFGKTAFD ISIDNGNEDI AEVLQKAAKL
K10 008-801-2178-C11 QDKWGTTPFD LAALIGNEDI AEVLQKAAKL
K11 008-801-2178-E11 QDHMGETPFD LAAYFGNEDI AEVLQKAAKL
K12 008-801-2179-E4 QDKFGKTPFD LAIDNGNEDI AEVLQKAAKL
K15 008-801-2179-E12 QDLRGQTPFD LAAWQGNEDI AEVLQKAAKL

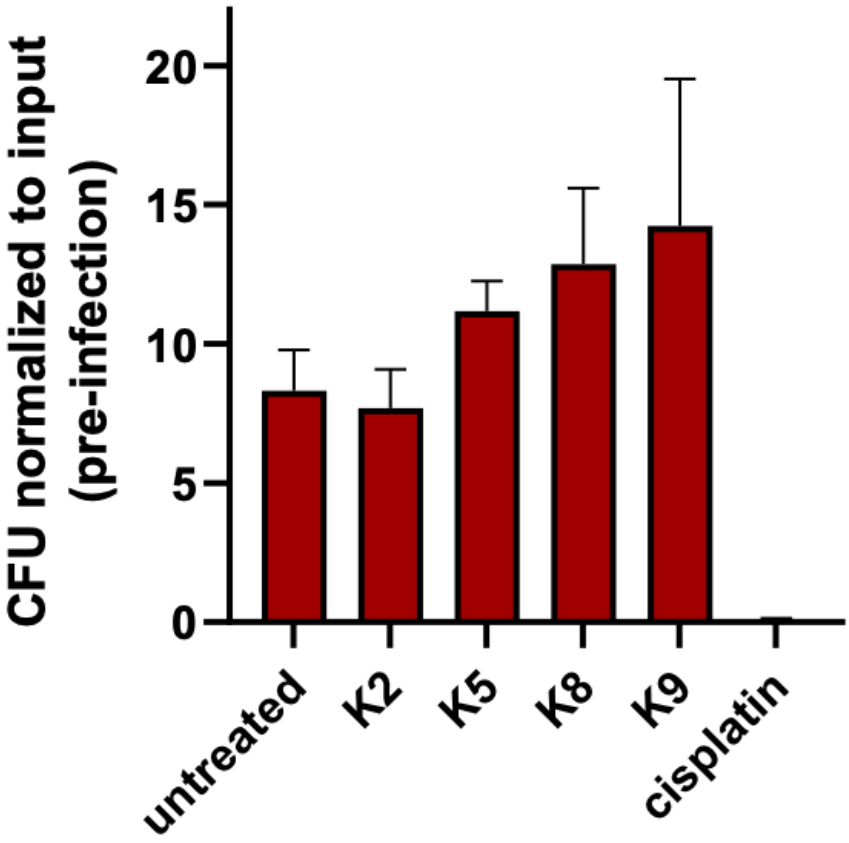
S6 Figure



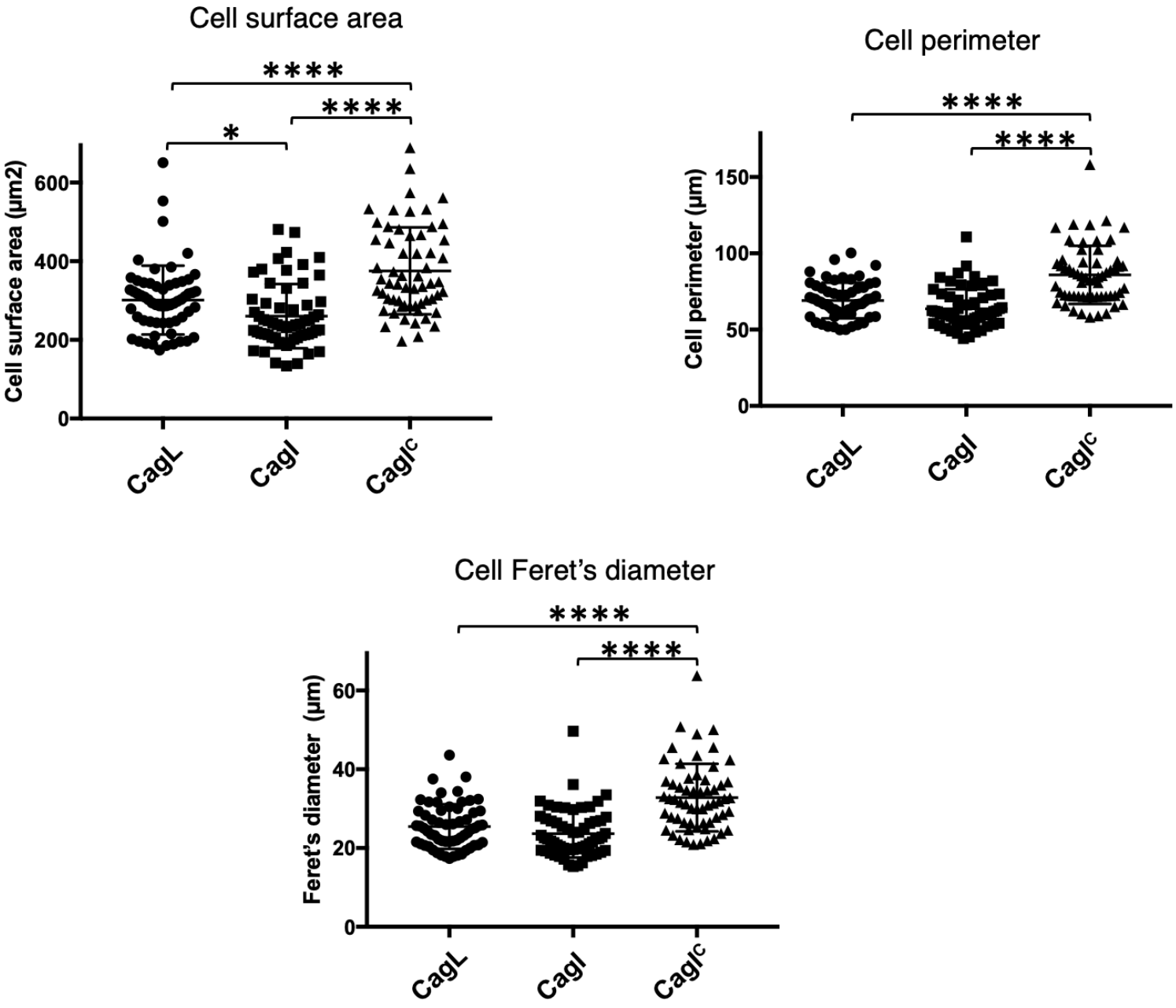
S7 Figure


CagI-K15
agregates
CagI-K10
95 kDa
CagI-K9
117 kDa
CagI-K8
95 kDa
CagI-K11
90 kDa
CagI-K12
123 kDa
UV Absorbance (UA)
Molar Mass (g/mol)
Volume (mL)

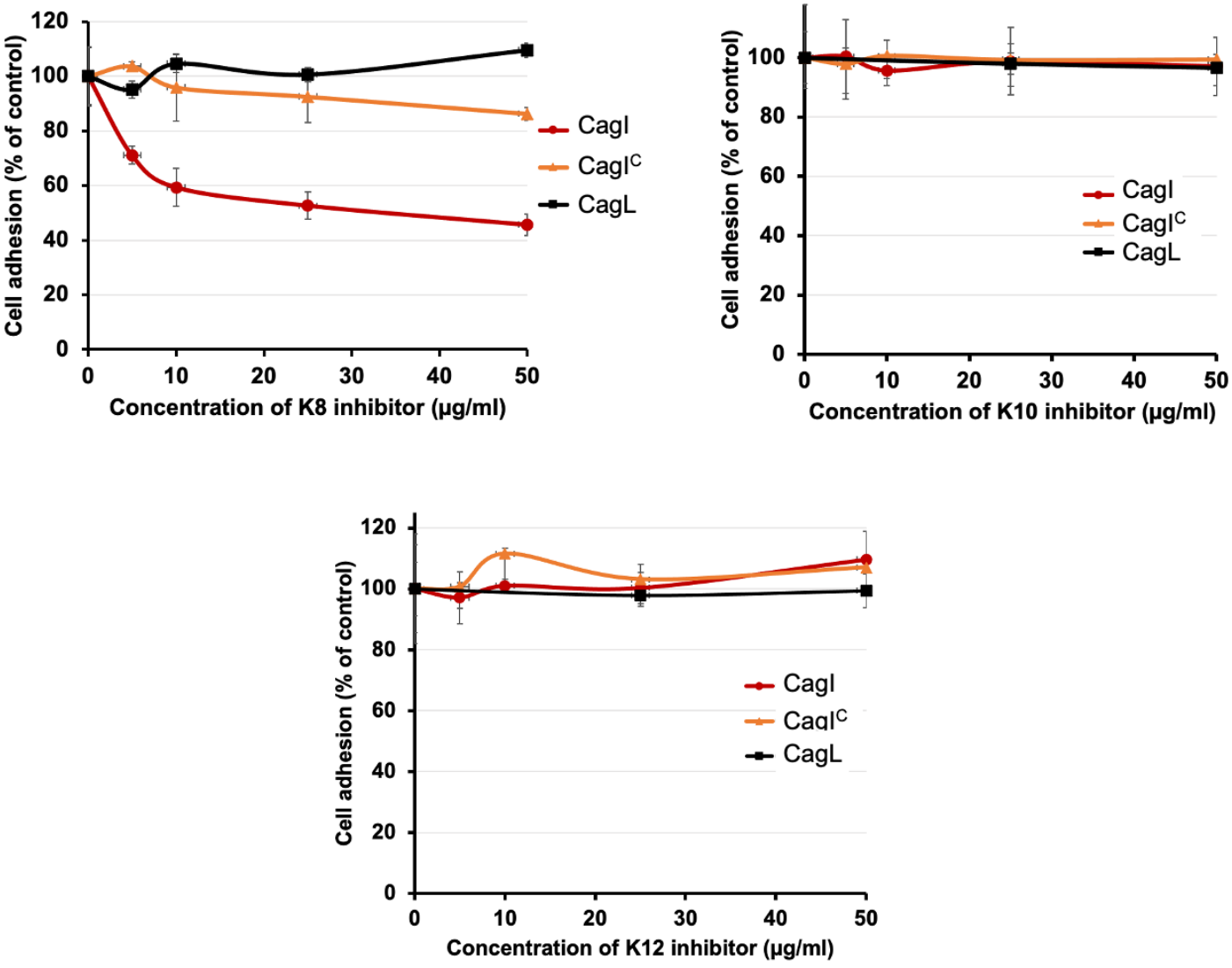
S8 Figure


CFU normalized to input
(pre-infection)
20
15
10
5
0
untreated
K2
K5
K8
K9
cisplatin

S9 Figure


Cell surface area
Cell surface area (µm2)
0
200
400
600
****
****
*
CagL
CagI
CagIc
Cell perimeter
Cell perimeter (µm)
0
50
100
150
****
****
CagL
CagI
CagIc
Cell Feret's diameter
Feret's diameter (µm)
0
20
40
60
****
****
CagL
CagI
CagIc

S10 Figure


Cell adhesion (% of control)
120
100
80
60
40
20
0
0
10
20
30
40
50
Concentration of K8 inhibitor (μg/ml)
CagI
CagI^C
CagL
Cell adhesion (% of control)
100
80
60
40
20
0
0
10
20
30
40
50
Concentration of K10 inhibitor (μg/ml)
CagI
CagI^C
CagL
Cell adhesion (% of control)
120
100
80
60
40
20
0
0
10
20
30
40
50
Concentration of K12 inhibitor (μg/ml)
CagI
CagI^C
CagL