The FEBS Journal

FEBS PRESS science publishing by scientists

Molecular dissection of protein–protein interactions between integrin α5β1 and the *Helicobacter pylori* Cag type IV secretion system

Thomas Koelblen[1,#], Célia Bergé[1,#], Mickaël V. Cherrier[1,†], Karl Brillet[1], Luisa Jimenez-Soto[2], Lionel Ballut[1], Junichi Takagi[3], Roland Montserret[1], Patricia Rousselle[4], Wolfgang Fischer[2], Rainer Haas[2,5], Rémi Fronzes[6] and Laurent Terradot[1]

1 UMR 5086 Molecular Microbiology and Structural Biochemistry, Institut de Biologie et Chimie des Protéines, CNRS-Université de Lyon, France
2 Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie, Ludwig-Maximilians-Universität, München, Germany
3 Laboratory of Protein Synthesis and Expression, Institute for Protein Research, Osaka University, Japan
4 Laboratoire de Biologie Tissulaire et Ingénierie Thérapeutique, UMR 5305, CNRS, University Lyon 1, France
5 German Center for Infection Research (DZIF), Partner Site LMU, München, Germany
6 European Institute of Chemistry and Biology, CNRS, UMR 5234, Microbiologie Fondamentale et Pathogénicité, University of Bordeaux, Pessac, France

Keywords
host–pathogen interaction; pilus; stomach cancer; surface plasmon resonance; type IV secretion system

Correspondence
L. Terradot, UMR 5086 Molecular Microbiology and Structural Biochemistry CNRS-Université de Lyon, Institut de Biologie et Chimie des Protéines, 7 Passage du Vercors, F-69367, Lyon Cedex 07, France
Fax: + 33 4 72 72 26 04
Tel: +33 4 72 72 26 52
E-mail: laurent.terradot@ibcp.fr

†**Present address**
CEA, CNRS, IBS, University of Grenoble Alpes, France

#These authors contributed equally to this work.

(Received 17 August 2017, revised 4 October 2017, accepted 17 October 2017)

doi:10.1111/febs.14299

The more severe strains of the bacterial human pathogen *Helicobacter pylori* produce a type IV secretion system (*cag*T4SS) to inject the oncoprotein cytotoxin-associated gene A (CagA) into gastric cells. This syringe-like molecular apparatus is prolonged by an external pilus that exploits integrins as receptors to mediate the injection of CagA. The molecular determinants of the interaction of the *cag*T4SS pilus with the integrin ectodomain are still poorly understood. In this study, we have used surface plasmon resonance (SPR) to generate a comprehensive analysis of the protein–protein interactions between purified CagA, CagL, CagI, CagY repeat domain II ($CagY^{RRII}$), CagY C-terminal domain ($CagY^{B10}$) and integrin α5β1 ectodomain ($\alpha5\beta1^{E}$) or headpiece domain ($\alpha5\beta1^{HP}$). We found that CagI, CagA, CagL and $CagY^{B10}$ but not $CagY^{RRII}$ were able to interact with $\alpha5\beta1^{E}$ with affinities similar to the one observed for $\alpha5\beta1^{E}$ interaction with its physiological ligand fibronectin. We further showed that integrin activation and its associated conformational change increased CagA, CagL and $CagY^{B10}$ affinities for the receptor. Furthermore, CagI did not interact with integrin unless the receptor was in open conformation. CagI, CagA but not CagL and $CagY^{B10}$ interacted with the $\alpha5\beta1^{HP}$. Our SPR study also revealed novel interactions between CagA and CagL, CagA and $CagY^{B10}$, and CagA and CagI. Altogether, our data map the network of interactions between host-cell α5β1 integrin and the *cag*T4SS proteins and suggest that activation of the receptor promotes interactions with the secretion apparatus and possibly CagA injection.

Abbreviations
Cag, cytotoxin-associated gene; FAK, focal adhesion kinase; K_D, dissociation constant; MALS, multi-angle light scattering; R_g, radius of gyration; RU, resonance unit; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; SLB, single-layer beta sheet; SPR, surface plasmon resonance; T4SS, type IV secretion system; TEV, tobacco etch virus.

Introduction

Type IV secretion systems (T4SSs) are bacterial molecular machineries that span both membranes and are prolonged outside the cell by a pilus. These apparatuses transport macromolecules across membranes, and many are found in bacterial pathogens, in which they are used to adhere and/or to translocate effectors into host cells to promote infection [1]. The genomes of the more severe strains of the human pathogen *Helicobacter pylori* contain the cytotoxin-associated gene (*cag*) pathogenicity island (*cag*PAI) that encodes for 28–30 proteins (reviewed in [2]). These proteins assemble a T4SS (named the *cag*T4SS) required to deliver the CagA oncoprotein into the host cell [3]. After translocation, CagA is tyrosine phosphorylated by host Src kinases and hijacks the signalling system of the cell [4]. This leads to morphological changes in the cell and uncontrolled proliferation and provokes tumour development [5].

The mechanism of injection of CagA by the *cag*T4SS is still poorly understood. The *cag*T4SS utilises integrin(s) as a host-cell receptor prior to substrate translocation (reviewed in [6]). Integrins form a family of heterodimeric mammalian cell surface receptors consisting of α and β subunits. Each integrin has three domains: an extracellular, transmembrane and cytoplasmic domain [7]. The extracellular parts of α and β subunits assemble into an ectodomain formed by a 'headpiece' and a 'tailpiece' domain (Fig. 1A). The β1 integrin family constitutes one of the largest subclass of integrins and is often associated with α5 integrin. Integrin is in equilibrium between low- and high-affinity states for ligands [8]. In the inactive (or resting) state, α5β1 adopts a bent-closed conformation [9]. Signals from the inside of the cell (inside-out signalling) can activate integrins, and the ectodomain undergoes a spectacular conformational change that results in an extended-closed conformation [7]. Further activation leads to headpiece opening to into the extended-open conformation that characterises a high-affinity state of the α5β1 receptor. The latter conformation is thought to be the most active for cell adhesion. Because of their ubiquitous distribution, β1

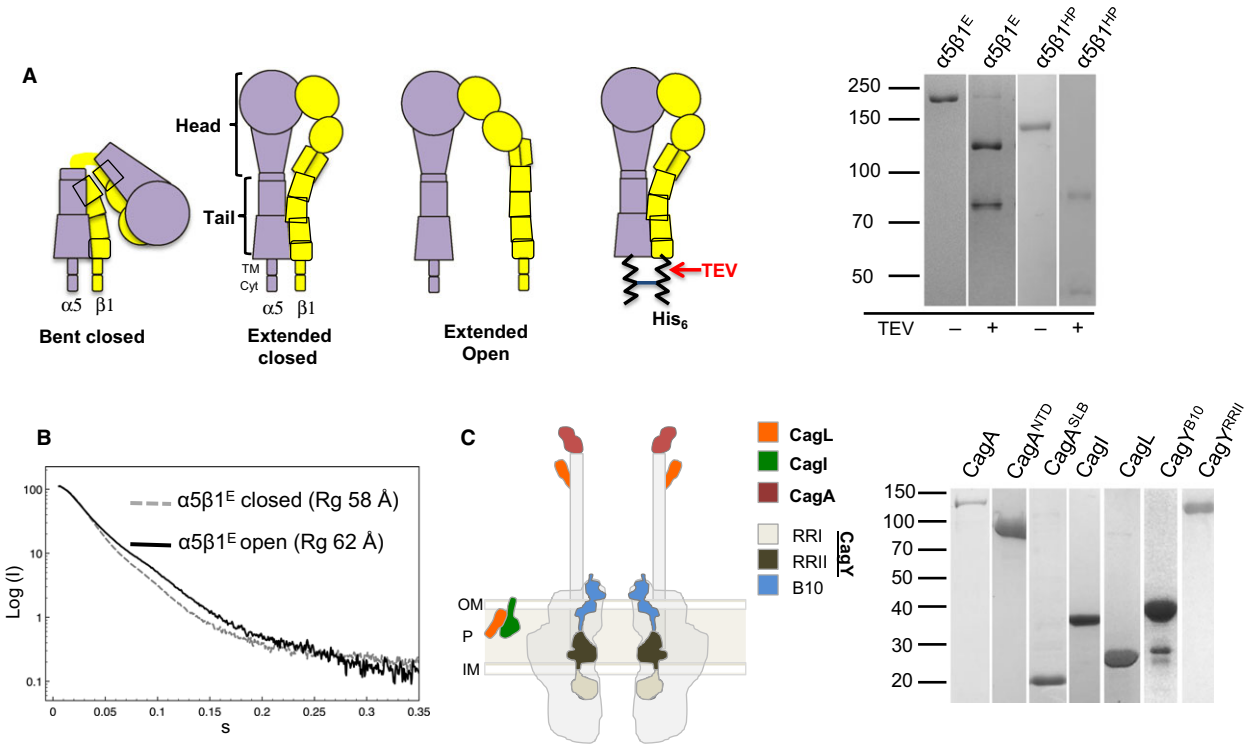

Fig. 1. (A) Schematic view of the three possible conformation of the integrin α5β1 with headpiece and tailpiece domain motion. The last α5β1 diagram shows the clasp system in the $\alpha5\beta1^{E}$ construct used in this study. Coomassie blue-stained SDS/PAGE of the purified ectodomain ($\alpha5\beta1^{E}$) or headpiece ($\alpha5\beta1^{HP}$) of α5β1 before and after TEV cleavage; molecular weights of the ladder are indicated on the left in kDa. (B) Experimental SEC-SAXS curves of $\alpha5\beta1^{E}$ before (black line) and after (grey dash) TEV cleavage. (C) Schematic view of the localisation of the putative integrin-interacting Cag proteins in the *cag*T4SS apparatus. OM, outer membrane; P, periplasm; IM, inner membrane. Coomassie blue-stained SDS/PAGE of the purified Cag proteins; molecular weights of the ladder are indicated in kDa on the left.

integrins are targeted by a large range of microbes that utilise them as receptors to promote adhesion, host-cell invasion or injection of virulence effectors [10–13].

Several proteins of the *cag*T4SS were found to interact with integrins (reviewed in [6]). The binding of the *cag*-pilus-associated CagL to the host-cell integrin activates the focal adhesion kinase (FAK) and Src tyrosine kinase, hereby activating CagA phosphorylation [14]. CagL contains the canonical integrin-binding motif arginine/glycine/aspartate (RGD), but there is no uniform agreement about the functional significance of this RGD motif in mediating binding of *H. pylori* to integrin [15]. Recombinantly expressed CagL functionally mimics fibronectin in its ability to induce focal adhesion formation in mouse fibroblasts and stimulate spreading of AGS cells [16] and mediates dissociation of ADAM17 from α5β1 [17]. It has been shown by surface plasmon resonance (SPR) that CagL has a higher affinity for the integrin than its RGD mutant $CagL^{RGA}$. CagL was also found to interact with integrins αVβ6, αVβ3 and αVβ8, but in an RGD-dependent manner [14,18–20]. However, CagL and $CagL^{RGA}$ mutant can bind to αVβ5 with a similar affinity, suggesting that the RGD motif is not always required for integrin binding [18].

The effector protein CagA and the *cag*T4SS proteins CagY and CagI were also found to interact with β1 using yeast two-hybrid and co-immunoprecipitation assays [21]. CagY has been detected at the surface of pilus-like structure [22], and CagI is also present at the surface of the bacteria [23]. CagI and CagL interact together [24] and together with CagH play a key role in cagT4SS pilus assembly [25].

Mechanistically, the interaction of CagA with integrin is intriguing because it suggests that the protein could participate actively to its translocation [21,26]. Along these lines, CagA is produced in large amounts, but relatively small amounts are injected into the cell, suggesting that it might have other functions in *H. pylori* [27]. The crystal structure of CagA N-terminal domain and the associated translocation competition experiments suggested that a single-layer beta sheet (SLB) domain of CagA is particularly important for the interaction with integrin [26]. *In vivo* studies also showed that integrin activation and clustering was involved in CagA injection by the *cag*T4SS [21]. Interestingly, locking integrin in its activated conformation on the AGS cells by targeted antibody blocked CagA translocation by *H. pylori* [21]. This suggests that structural changes of integrin are important for the translocation mechanism of the oncoprotein.

To better characterise the molecular interplay between integrin α5β1 and Cag proteins, we set out to purify all known and putative binding partners of the receptor and performed a comprehensive SPR study. We used a molecular system to manipulate integrin α5β1 conformation in order to determine whether the latter influences the interactions with Cag proteins. Our study provides a molecular mapping of the interactions involved in *cag*T4SS interplay with its receptor that enables the recruitment and clustering of integrins associated with CagA injection.

Results

Protein purification and characterisation

To purify integrin α5β1 ectodomain, we used two different constructs: one containing both the tailpiece and headpiece domains ($α5β1^{E}$, ectodomain) and another one containing only the headpiece ($α5β1^{HP}$, headpiece). The α5β1 proteins produced contain a His-tag and an ACID (α5)–BASE (β1) molecular 'clasp' as described previously in [28,29]. As a consequence, the purified integrins are joined by this clasp at the C terminus, which simulate a coiled-coil formed by the membrane and cytoplasmic-tail domains. This clasp device allows for selective activation of the integrin by the tobacco etch virus (TEV) protease. The purified integrins adopt a conformation with tails joined together by the molecular clasp but with a headpiece domain in a closed state. A recent structural study on a similar construct showed that in this condition, $α5β1^{E}$ adopts either a bent or extended, or a range of intermediate conformations with the headpiece closed [9]. When the clasp is removed by digesting the sample with the TEV protease, the tail domains are separated resulting in an opening of the headpiece and a complete extended conformation [29] (Fig. 1A). To ascertain the conformational change between the two proteins samples, size-exclusion chromatography (SEC) coupled to small-angle X-ray scattering (SAXS) experiments was performed (Table 1). As shown in Fig. 1B, the structural change associated with TEV cleavage is important as the two heterocomplexes have different SAXS curves in solution. This structural change was also supported by significant change in the radius of gyration (R_g) calculated from the two samples that was estimated to be around 58 Å and 62 Å before and after TEV cleavage, respectively (Table 1).

Cag proteins and/or domains were purified recombinantly to near homogeneity (Fig. 1). CagA was purified as full-length or as fragments comprising residues 1 to 884 ($CagA^{NTD}$) or 303 to 456 ($CagA^{SLB}$) as previously described [26] (Fig. 1). CagI and CagL were expressed as full-length proteins without their

Table 1. SAXS data collected and analysis performed on integrin $\alpha5\beta1^{E}$ before (closed) and after (open) TEV cleavage of the molecular clasp.

	$\alpha5\beta1^{E}$ closed	$\alpha5\beta1^{E}$ open
Data collection		
Beamline	SOLEIL SWING	
Wavelength (Å)	1.003	
q range ($Å^{-1}$)	0.002–0.351	
Detector	PCCD170170 (AVIEX)	
Detector distance (m)	1.8	
Injection volume (µL)	50	
Temperature (K)	288	
Structural parameters		
R_g (Å) [from $P(r)$]	58.1	62.5
R_g (Å) [from Guinier]	57 ± 0.2	59.1 ± 0.2
D_{max} (Å)	225	241
Porod volume estimate ($Å^3$)	347 000	211 000
Molecular mass calculated	224	168
Molecular mass predicted from sequence	189	189

predicted signal peptides (Fig. 2). CagY is a multidomain protein of 250 kDa (strain 26695) containing two repeat regions, repeat region I (RRI) and II ($CagY^{RRI}$ and $CagY^{RRII}$, respectively), and a C-terminal domain ($CagY^{B10}$) having some sequence similarity with VirB10 from *Agrobacterium tumefaciens* (Fig. 3) [30]. The production of full-length CagY proved difficult, so we purified $CagY^{RRII}$ (residues: 464–1460, strain 26695) that was previously detected on the pilus [22] and $CagY^{B10}$ (residues: 1542–1896, strain P12) possibly located at the outer membrane (OM) of the bacteria based on the structure of the *Escherichia coli* T4SS core complex [31]. The oligomeric states of purified CagL, CagI, $CagY^{B10}$ and $CagY^{RRII}$ were investigated by SEC coupled to multi-angle light scattering (SEC-MALS), and the secondary structure properties were studied by CD. CagL secondary structure content corresponded well to the one derived from the crystal structure (Fig. 2) [32]. The protein was purified as a monomer, but presence of dimers was detected in SEC-MALS as previously suggested (Fig. 2) [32]. CagI was for the first time purified here, and CD indicated that the protein is predominantly helical (Fig. 2). The protein was able to form different oligomers (trimers or higher) as determined by SEC-MALS. As shown in Fig. 3, $CagY^{B10}$ secondary structure content corresponded well to the predicted modified β-barrel based on crystal structure of its homologue ComB10 [33]. The $CagY^{RRII}$ contains almost exclusively α-helices as previously found by secondary structure predictions and CD analysis [30]. Both $CagY^{B10}$ and $CagY^{RRII}$

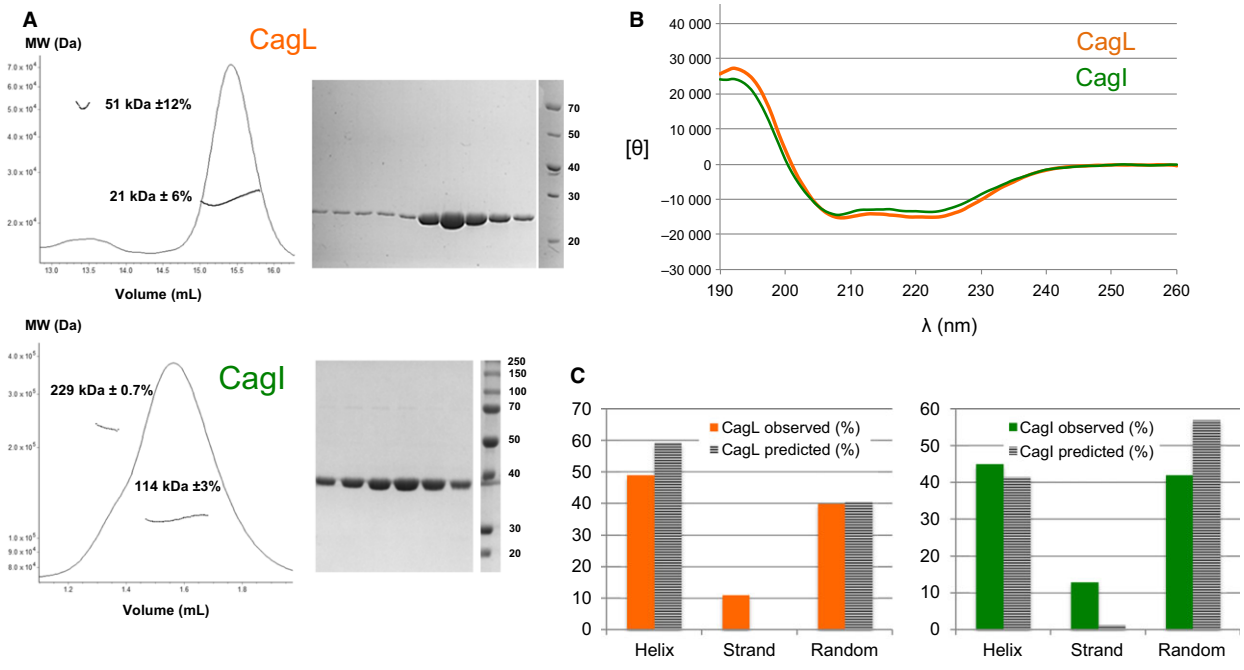

Fig. 2. (A) SEC-MALS analysis of CagL and CagI and corresponding SDS/PAGE analysis of the elution peak fractions. (B) CD measurements for CagL and CagI. (C) Comparison of secondary structure content determined experimentally and predicted by *in silico* analysis, PSI pred for CagI and the crystal structure for CagL (PDB code 3ZCJ).

were monomeric in solution as shown by SEC-MALS (Fig. 3).

Integrin α5β1 ectodomain interactions with Cag proteins

We investigated the interactions between Cag proteins and integrin by performing SPR experiments. All experiments were performed in both directions, and specificity of binding was assessed using BSA and fibronectin as negative and positive control, respectively. No interactions were detected with BSA either immobilised or injected as control. We observed that when integrin ectodomain was immobilised, binding of Cag proteins or fibronectin was detected. Interactions generally were stronger with the open form of $\alpha5\beta1^{E}$ than with the closed form (Fig. 4). However, several of these experiments could not be reproduced consistently. This was likely due to issues in regenerating the immobilised sensorchip and a possible heterogeneity in integrin conformation on the sensorchip surface. Similar issues were observed using commercially available full-length integrin α5β1, α1β1 and αVβ3 (data not shown). This suggests that immobilising the integrin was not the best strategy for studying these interactions.

When Cag protein or fibronectin was immobilised, the binding experiments were reliable (Fig. 5). The dissociation constant (K_D) values obtained are indicated in Table 2. All proteins except $CagY^{RRII}$ and CagI could interact with integrin $\alpha5\beta1^{E}$ in the closed conformation. The signals obtained with $CagA^{SLB}$ were very low [< 5–9 resonance unit (RU); Fig. 5], and thus, the K_D was not reliable. CagA, $CagA^{NTD}$, $CagY^{B10}$ and CagL interacted with integrin with affinities in the range of hundred nanomolar (Table 2). These K_D values were similar to the ones obtained with fibronectin (Table 2). In previous SPR studies, CagL was found to interact with commercial full-length αVβ5 with a K_D of 200 nM [18] and α5β1 with a K_D of 90 nM [14], thus with affinities comparable to the values obtained here. In a previous study, we found that $CagA^{NTD}$

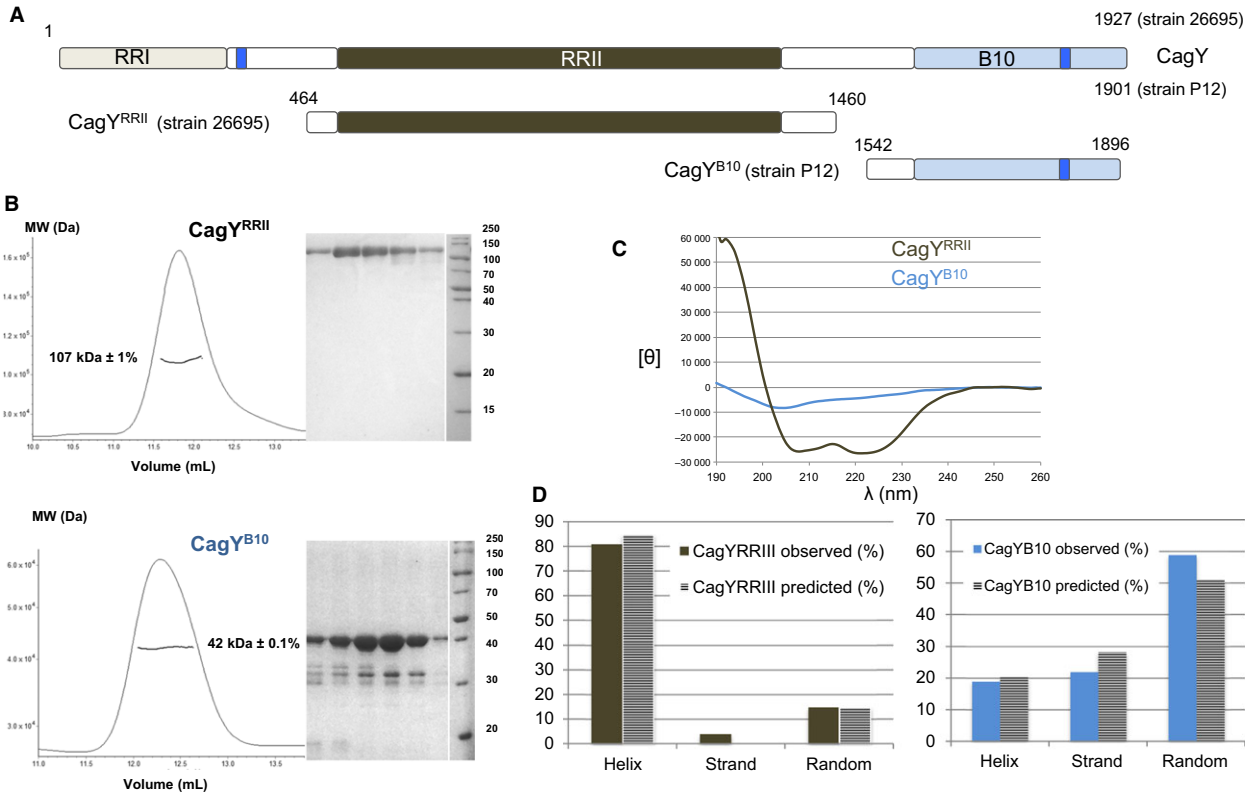

Fig. 3. (A) Schematic representation of CagY domain organisation with RRI and RRII and the predicted transmembrane helices (blue boxes). The length of the protein in amino acids is indicated for CagY of *Helicobacter pylori* strain P12 or 26695. Below are represented the two constructs used in this study. (B) SEC-MALS analysis of $CagY^{B10}$ and $CagY^{RRII}$ and corresponding SDS/PAGE analysis of the elution peak fractions. (C) CD measurements for of $CagY^{B10}$ and $CagY^{RRII}$. (D) Comparison of secondary structure content determined experimentally and predicted by *in silico* analysis: PSI pred (http://bioinf.cs.ucl.ac.uk/psipred/) for $CagY^{RRII}$ and the crystal structure of ComB10 (PDB code 2BHV).

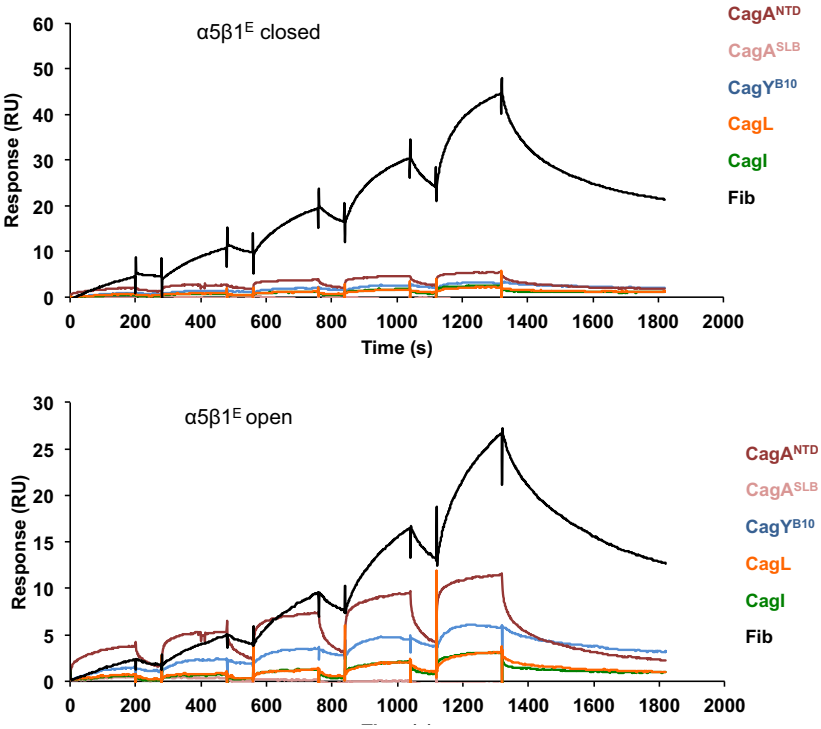

Fig. 4. Single-injection-mode experiments using CM5 chips coated with integrin ectodomain ($\alpha5\beta1^E$) before or after TEV cleavage. Cag proteins were injected on the chips at increasing concentrations.

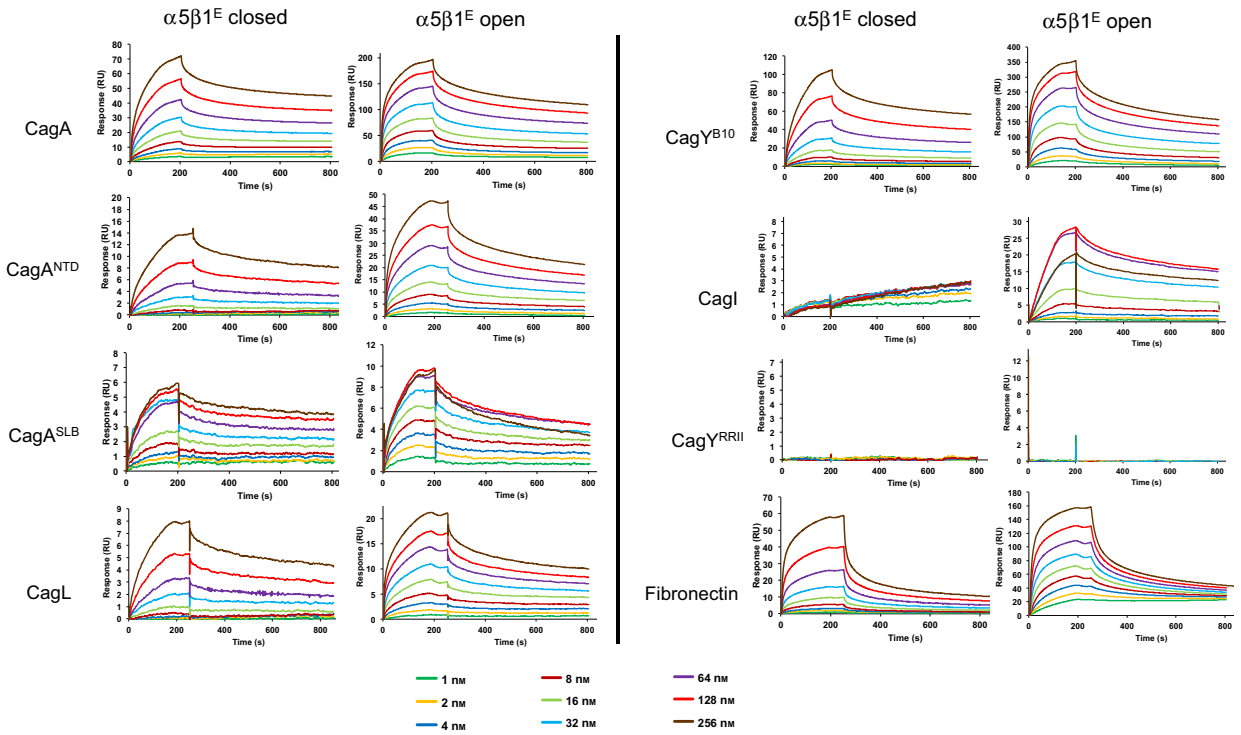

Fig. 5. Comparison of the kinetics of Cag protein binding to recombinant integrin α5β1 ectodomain ($\alpha5\beta1^E$) before (closed) and after (open) TEV cleavage. Cag proteins or fragments were immobilised on the sensorchip at the same density (1000 RU), and integrin was injected at 20 $\mu L \cdot min^{-1}$. Traces show increasing concentrations (1–256 nM; corresponding values are indicated on the graph) of analyte integrin.

Table 2. SPR affinity values expressed in nM (average of three replicates). Interactions were measured between pilus-associated Cag proteins CagA, $CagA^{NTD}$, CagA, $CagA^{SLB}$, CagL, CagI, $CagY^{RRII}$ and $CagY^{B10}$ proteins and integrin α5β1 ectodomain ($\alpha5\beta1^{E}$) or fibronectin. Integrins were either clasped or unclasped to mimic closed and open states, respectively. Cag proteins and fibronectin were immobilised on a CM5 sensorchip. –, no interaction; nd, K_D not determined.

Cag protein	$\alpha5\beta1^{E}$ closed	$\alpha5\beta1^{E}$ open
CagA	126 ± 32	30 ± 4
$CagA^{NTD}$	226 ± 109	56 ± 24
$CagA^{SLB}$	nd	153 ± 34
CagL	183 ± 97	39 ± 20
CagI	–	15 ± 2
$CagY^{B10}$	237 ± 67	41 ± 8
$CagY^{RII}$	–	–
Fibronectin	167 ± 87	39 ± 21

interacted with full-length α5β1 with a much higher affinity (K_D of 0.15 nM) [21]. Such differences could be due to the differences in the constructs used. It is possible that the commercial integrin adopts a conformation that has a higher affinity for $CagA^{NTD}$ binding or that $CagA^{NTD}$ binds to additional domains of integrin, that is the transmembrane domain (TM) and the cytoplasmic tail (CT). Alternatively, this difference might be due to glycosylation of the ectodomain of commercial α5β1 that was purified after extraction from human tissue.

When integrin was switched to the open conformation, affinities were increased nearly 4-fold with K_D values of 30 nM (CagA), 56 nM ($CagA^{NTD}$), 39 nM (CagL) and 41 nM ($CagY^{B10}$) (Table 2). A strong interaction was observed with CagI with a K_D of 15 nM. Interaction was observed with $CagA^{SLB}$ but with a higher K_D (153 nM), and no interaction was observed with $CagY^{RRII}$. These results show that the CagA, CagI, CagL and $CagY^{B10}$ proteins interact with high affinity with integrin $\alpha5\beta1^{E}$ and that activation of integrin significantly increases the affinities of the Cag proteins/integrin interaction. Furthermore, the results confirm that the C-terminal part of CagA is not required for CagA interaction and the SLB domain interacts directly with integrin as previously proposed [26].

Interactions of α5β1 headpiece domain with Cag proteins

To better map the interacting site of integrin with the Cag proteins, we used a fragment of α5β1 that contains only the headpiece domain ($\alpha5\beta1^{HP}$). Cag proteins were immobilised, and headpiece fragments were injected on the CM5 sensorchips (Fig. 6). Under these conditions, we observed only an interaction of CagA and fibronectin with $\alpha5\beta1^{HP}$ closed. As shown in Table 3, the K_D of $CagA^{NTD}$ with $\alpha5\beta1^{HP}$ was similar to the one obtained with $\alpha5\beta1^{E}$: 143 nM and 226 nM, respectively.

When $\alpha5\beta1^{HP}$ conformation was switched to open, the K_D of the interaction with $CagA^{NTD}$ was around 30-fold stronger (5 nM) than with the closed conformation. $CagA^{SLB}$ was found to interact with the headpiece with a K_D value of around 66 nM but only in the open conformation (Fig. 6; Table 3). CagI did not interact with the closed form of the integrin headpiece, but a clear interaction was observed with $\alpha5\beta1^{HP}$ in

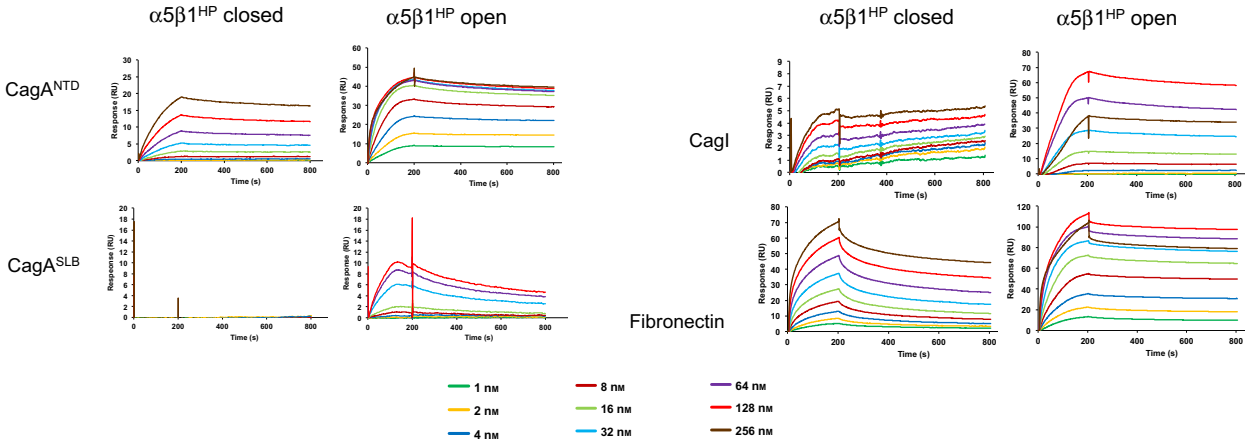

Fig. 6. Kinetics of Cag protein binding to recombinant integrin α5β1 headpiece domain ($\alpha5\beta1^{HP}$) before (closed) and after (open) TEV cleavage. Cag proteins or fragments were immobilised on the sensorchip at the same density (1000 RU), and integrin was injected at 20 μL·min^{-1}.

Table 3. K_D values expressed in nM. Interactions were measured between pilus-associated proteins CagA, $CagA^{NTD}$, $CagA^{SBL}$, CagL, CagI, $CagY^{RRII}$ and $CagY^{B10}$ proteins and integrin α5β1 headpiece domain ($\alpha5\beta1^{HP}$). Integrins were either clasped or unclasped to mimic closed and open states, respectively. Cag proteins and fibronectin were immobilised on a CM5 sensorchip.

Cag protein	$\alpha5\beta1^{HP}$ closed	$\alpha5\beta1^{HP}$ open
$CagA^{NTD}$	143[a]	5 ± 2
$CagA^{SLB}$	–	66 ± 12
CagL	–	–
CagI	–	53 ± 24
$CagY^{B10}$	–	–
$CagY^{RRII}$	–	–
Fibronectin	40[a]	9[a]

[a]A K_D was derived from a single experiment.

open conformation (K_D of 53 nM; Table 3). No interaction was observed with either CagL or $CagY^{B10}$ under these conditions.

Protein interactions between Cag-pilus-associated proteins

CagA and CagL were detected at the tip of the *cag*T4SS pilus [14,21]. It was previously found that CagI and CagL can interact *in vivo* [24]. Using a similar SPR set-up, we investigated the interactions between Cag proteins that interacted with integrins (Fig. 4 and Table 4). As a standard, we used the chaperone CagF for which detailed interaction studies with CagA have been performed *in vivo* [34,35] and *in vitro* [36]. In our conditions, CagF interacted with full-length CagA with a K_D of 51 nM (Fig. 7, Table 4), in agreement with previous isothermal titration calorimetry studies that found that CagF binds to CagA with a K_D of 49 nM [36]. CagA interacted with $CagY^{B10}$ in a similar affinity range than with CagF (K_D of 47 nM) and with a lower affinity with CagL and CagI (K_D of 213 and 224 nM, respectively; Table 4). The N-terminal domain of CagA was also found to interact with $CagY^{B10}$, CagL and CagI with higher affinity (K_D of 76, 70 and 55 nM, respectively) than with CagF (K_D of 183 nM). An interaction was also observed between CagI and CagL and between $CagY^{B10}$ and CagL (Fig. 7). However, the K_D could not be reliably determined by our SPR experiments due to lack of binding saturation and low signal.

Table 4. SPR affinity values expressed in nM (average of three replicates). – indicates no interaction. nd indicates interaction detected but the K_D not determined.

	Immobilised			
	CagL	CagI	$CagY^{B10}$	CagF
Injected				
CagA	213 ± 40	224 ± 50	47 ± 12	51 ± 7
$CagA^{NTD}$	70 ± 17	55 ± 23	76 ± 5	183 ± 51
CagL	nd	nd	nd	–
CagI	nd	–	–	–
$CagY^{B10}$	nd	–	24 ± 8	–

Discussion

The *H. pylori cag*T4SS is a molecular machine that uses integrin α5β1 as a receptor to inject the oncoprotein CagA into gastric epithelial cells. Four proteins of the *cag*T4SS were found to interact with the receptor: CagL, CagA, CagY and CagI [14,21,26]. The proteins CagL and CagA have been particularly well studied, and direct interactions have been measured with purified proteins. However, only few studies have been performed with purified CagY and CagI. Moreover, very little was known about the domain of the integrin α5β1 involved in these interactions and about the influence of integrin α5β1 conformation on the binding. Here, we have purified the proteins and investigated the protein–protein interaction between integrin α5β1 and the pilus-associated Cag proteins from *H. pylori* using SPR.

Our work demonstrates that CagA, CagL, CagY and CagI directly interact with integrin α5β1 ectodomain and hereby confirm previous studies [21]. Interestingly, we found that CagA and CagI bind to the headpiece domain of the receptor. Given that none of these proteins display a RGD motif, the binding to the headpiece by these two proteins involves a different motif. In the case of CagA, our data support that the SLB domain interacts directly with integrin headpiece, but other domains of $CagA^{NTD}$ might also participate. This is consistent with our previous proposal that the SLB is an integrin-binding motif [26]. CagI is likely to interact with integrin via a different mode than CagA, given that the protein is principally composed of α-helices as found by CD.

CagL was found to interact with the ectodomain of integrin but not with the headpiece domain. A large number of studies have shown that the CagL RGD motif was important for binding to integrins [14,16,19,32,37–39]. The structure of CagL has revealed that the RGD motif is part of an α-helix [32] and thus differs from the common RGD motifs that are usually found in extended loops [11]. However, it was proposed that at acidic pH, the conformation of the RGD-containing α-helix could change and become structurally available for integrin binding [38]. The

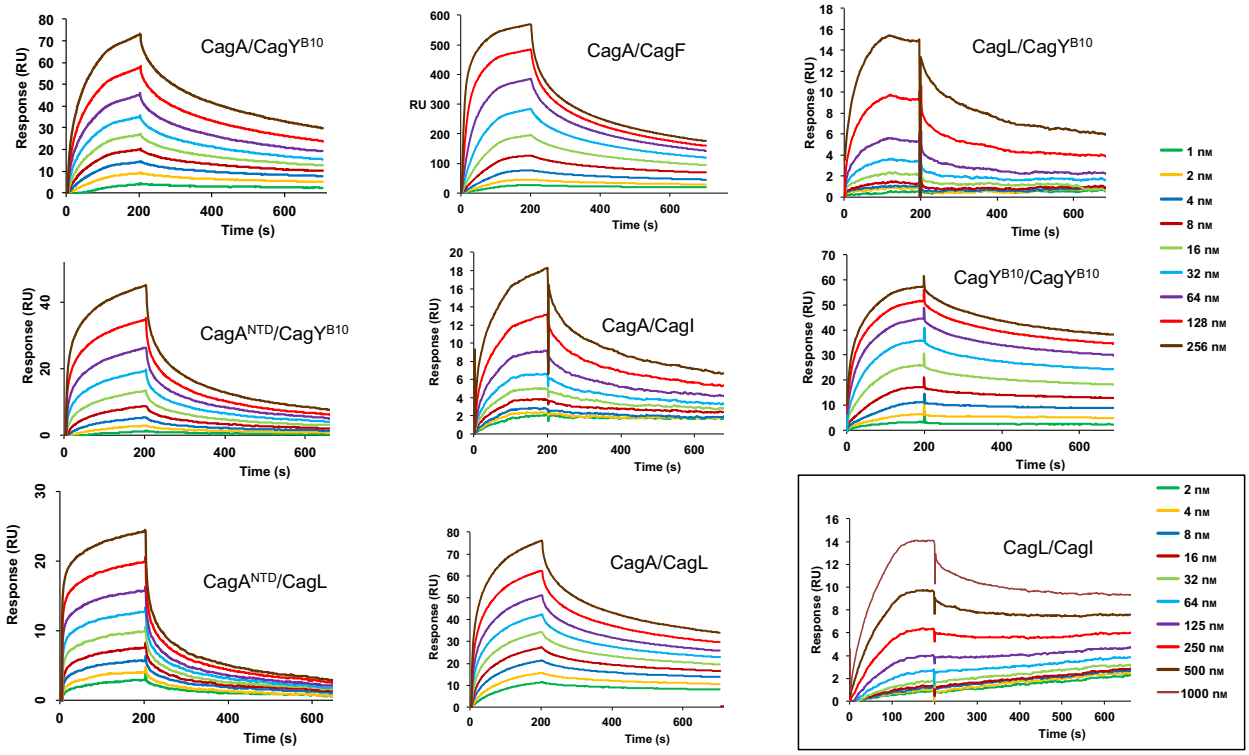

Fig. 7. Sensorgrams of Cag protein binding to Cag proteins or fragments. The sensorgrams correspond to the protein immobilised/analyte. The concentration of the analyte is indicated on the right except for the CagL/CagI interaction (inset).

crystal structure of this α5β1 headpiece domain revealed that the prototypical interaction of the RGD-containing proteins occurs at a groove located at the interface between the β-propeller (α5 subunit) and βI domain (β1 subunit) [28]. In our condition, fibronectin could interact efficiently with the headpiece domain. Our results thus suggest that the RGD motif of CagL does not bind to α5β1 in a 'canonical' mode. We cannot rule out that in the SPR experiments, the structure of CagL is affected and that the conformation of the helical RGD motif is not appropriate. However, this is unlikely as the protein was able to efficiently interact with the α5β1 ectodomain. Beside the RGD, several CagL motifs were found to be involved in the interaction with integrins. These include TSPSA and TASLI motifs that were recently identified in CagL to be important for binding to α5β1 [37]. The RGDLXXL motif of CagL is potentially involved in the interaction with αVβ6 [39]. This motif is found in TGF-β, a natural ligand of αVβ6 and interacts with the βI domain of the β6 subunit (part of the headpiece). Our results point towards an interaction of CagL with the tailpiece of integrin α5β1 rather than with the headpiece. Thus, it is likely that CagL binds in different ways to α5β1 and αVβ6. This would explain why the RGD motif was found not required for α5β1 binding in some studies [21], but further work will be required to determine the structural basis of CagL binding to specific integrins. Along these lines, we found that the CagI and CagL proteins interact with different domains of α5β1, and thus, a complex containing CagL and CagI complex could interact with very high affinity to the entire ectodomain.

In our SPR experiments, the C-terminal part of CagY, that is the domain sharing homology with the VirB10 family of proteins, was found to interact with the integrin ectodomain. VirB10 and its homologues are essential components of T4SS composed of a flexible α-helical N-terminal domain and a C-terminal modified β-barrel [33,40]. In the pKM101 T4SS assembly, TraF (VirB10 homologue) forms a large tetradecamer complex with TraO (VirB9) and TraN (VirB7) [31,41]. Most of TraF β-barrel is located in the OM, and only the loop belonging to the so-called antennae is accessible to the surface of the bacteria [31]. In the *cag*T4SS, the core complex seems to be conserved with CagT (VirB7 homologue), CagX (VirB9 homologue) and CagY [42], although the assembly is larger and

involves more proteins than the pKM101 one [43]. The loop accessible at the surface might represent a good candidate for interaction with integrin, although no known integrin-interaction motif could be identified in the molecule. From a mechanistic point of view, the core complex seems to be always present in *H. pylori* cells, while the pilus production is induced by its contact with host cells [22,43]. It is therefore possible that the interaction of CagY with integrin takes place, while the pilus is not yet assembled and that CagY is involved in the early steps of *cag*T4SS interactions with the host cell prior to engagement of the pilus-associated proteins CagL, CagI and CagA.

Our results also reveal that CagA interacts with CagL and CagI and with the VirB10 part of CagY. CagY also interacts with CagL, and CagL interacts with CagI. It is difficult to envision CagA as part of a multiprotein complex that would include CagL, CagI and CagY and then be subsequently delivered into the host cell. However, all these proteins have been detected at the surface of the bacteria, and thus, such complexes might exist at the surface or on the *cag*T4SS pilus [14,21,23]. That CagA can interact with other pilus proteins and with integrin and is also injected into the cell strongly suggests that different CagA molecules have different functions. Some might play a role during the assembly of the *cag*T4SS pilus/receptor complex, while others are effector molecules. Indeed, CagA is produced in large quantities, but very little is injected [27]. A pool of CagA might thus participate in the assembly of the translocation-competent pilus by assembling the pilus proteins around the receptor. Interestingly, CagA seems to be also important for the assembly of the *cag*T4SS core complex [43]. The interaction identified here between CagA and CagL could take place at the tip of pilus as they have both been detected there [14,21]. The interaction between CagA and CagL and the C-terminal part of CagY might reflect interactions between these pilus-associated proteins and the OM translocation channel (CagY) prior to or during translocation of these proteins to the pilus. These novel interactions between Cag proteins should be further studied to determine their roles in the assembly and function of the *cag*T4SS.

We found that the extended-open conformation of α5β1 increases the affinities of the interactions with the *cag*T4SS proteins in a similar manner than with fibronectin [21]. It is particularly significant in the case of CagI, for which no interaction with integrin was detected unless the receptor was in its activated conformation, that is extended-open. This could be due to the oligomeric state of CagI (trimer or higher), which would sterically hinder binding to integrin. Opening of the head domain of α5β1 and/or separation of the α-leg and β-leg could thus allow the oligomer of CagI to bind to the integrin. Alternatively, integrin activation might also be a selective mechanism for CagI binding at a given moment. Interestingly, the extended-open conformation of α5β1, but not the extended-closed nor bent-closed, was found to be required for cell adhesion [9]. The higher affinity of the *cag*T4SS proteins for activated integrin could thus explain why integrin activation by Mn^{2+} ions was found to promote injection of CagA by *H. pylori* in AGS cells [21]. Our work also reveals that not all the Cag-pilus-associated proteins seem to target the same domain of α5β1. CagA and CagI seem to bind to the head domain of α5β1, but not CagY and CagL. Thus, our work raises the possibility that some of these interactions might not be mutually exclusive and that a dynamic interaction network occurs at the pilus–receptor interface (Fig. 8). Altogether, our study and the methods described here pave the way to understand the exploitation of integrin by *H. pylori* and possibly other bacterial pathogens.

Materials and methods

Overexpression and purification of proteins

Integrin α5β1

Two Chinese hamster ovary (CHO) lec 3.2.8.1 stable cell lines expressing two different constructs of α5β1 were used: one expressing the complete ectodomain, for example both the tailpiece and headpiece domains ($\alpha5\beta1^{E}$), and another one expressing only the headpiece ($\alpha5\beta1^{HP}$). These constructs have been described previously [28,29,44]. For production of the $\alpha5\beta1^{E}$, the expression construct for the α5 subunit contained residues 1–954 followed by a 30-residue ACID-Cys peptide, and the construct for β1 contained residues 1–708 followed by a TEV protease cleavage site, a 30-residue BASE-Cys peptide and a His_6-tag. For the $\alpha5\beta1^{HP}$, similar constructs were used except that α5 encompassed only residues 1–623 and β1 residues 1–445.

$\alpha5\beta1^{E}$ and $\alpha5\beta1^{HP}$ were purified using the same protocol. Supernatant containing the integrins was submitted to ammonium sulphate precipitation at 50% to remove contaminants. After a 16 000 ***g*** centrifugation (20 min at 4 °C), the supernatant was submitted to ammonium sulphate precipitation at 60% to collect integrins. The pellet obtained after centrifugation (16 000 ***g***, 20 min at 4 °C) was resuspended in buffer A (20 mM Tris pH 8.0, 1 M NaCl, 20 mM imidazole, 1 mM $CaCl_2$, 1 mM $MgCl_2$) and dialysed overnight against the same buffer. The solution was loaded onto a HisTrap column (GE Healthcare), and the protein was eluted with a linear gradient of buffer B

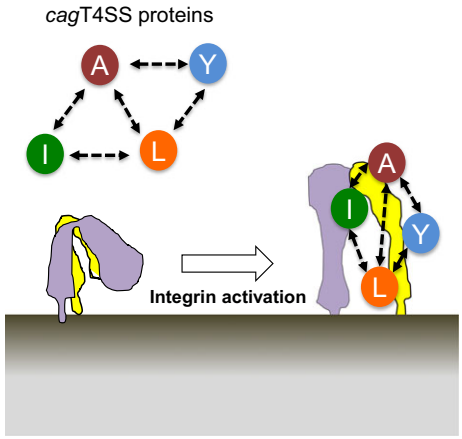

Fig. 8. Schematic representation of protein–protein interactions identified in this study. CagA, CagY, CagI, CagL and integrin proteins are coloured as in Fig. 1.

(20 mM Tris pH 8.0, 200 mM NaCl, 250 mM imidazole, 1 mM $CaCl_2$, 1 mM $MgCl_2$). Fractions were pooled and subjected to cleavage by the Endo H glycosidase. For integrin activation, the sample was then subjected to TEV cleavage overnight. The proteins were concentrated and loaded onto a Superdex 200 Increase 10/300 GL column equilibrated with 20 mM Tris pH 7.4, 137 mM NaCl, 2.7 mM KCl and 2% glycerol (v/v).

CagA

The *cagA* (strain P12) gene was inserted into the vector pETM30 (EMBL) between the *Nco*I and *Xho*I restriction sites. The sequence of *cagA* gene was mutated to remove the asparagine residues 880–885 (NNNNNN) by a single serine to reduce proteolysis occurring at this position during expression [45,46]. The resulting CagA protein was therefore expressed in fusion with a N-terminal His_6-GST-TEV- and a C-terminal STREP-tag. *Escherichia coli* BL21(DE3) strain carrying the plasmid was grown at 37 °C in LB medium with kanamycin (50 $g{\cdot}L^{-1}$), and protein expression was induced by 1 mM IPTG for 3 h at 30 °C. Purification of CagA was performed as in [47], and $CagA^{NTD}$ and $CagA^{SLB}$ were purified as previously described in [26].

CagL

The sequence encoding for CagL residues 21 to 237 (strain 26695) was inserted into the pRSF-MBP vector, a modified pRSF-Duet1 vector (Novagen) expressing the proteins in fusion with a N-terminal His_6-maltose binding protein (His_6-MBP) cleavable by TEV. *Escherichia coli* BL21 (DE3) strain carrying the vectors was grown in LB at 37 °C until the OD_{600} reached 0.6, and protein expression was induced for 16 h at 16 °C by adding 1 mM IPTG. Cells were harvested by centrifugation at 7000 ***g*** and resuspended in lysis buffer [50 mM Tris pH 8, 300 mM NaCl, 10% glycerol (v/v), 1% Triton (v/v)] with protease inhibitor tablet (complete EDTA-free; Roche), lysozyme (Roche) and DNase I (Sigma-Aldrich). The cells were lysed by sonication and centrifuged at 10 000 ***g*** for 15 min. The soluble fraction was separated from membrane fraction after centrifugation at 200 000 ***g*** for 45 min. The soluble fraction was loaded onto a HisTrap column (GE Healthcare). The protein was eluted with a 0–100% gradient of buffer (50 mM Tris pH 8, 200 mM NaCl, 5% glycerol, 500 mM imidazole). The His_6-MBP-tag of CagL was cleaved by the TEV protease during dialysis against 50 mM Tris pH 8, 200 mM NaCl and 5% glycerol. The dialysed protein was loaded onto a HisTrap 5-mL column and MBPTrap column (GE Healthcare). The cleaved proteins were collected in the flowthrough. The cleaved protein fractions were pooled, concentrated and loaded onto a Superdex 200 Increase 10/300 GL gel filtration column (GE Healthcare) equilibrated in 50 mM Tris pH 8, 200 mM NaCl and 5% glycerol.

CagI

CagI residues 21 to 381 (strain 26695) were inserted into the pRSF-MBP vector. BL21 (DE3) cells carrying His_6-MBP-CagI expression vector were grown in LB at 37 °C until an OD_{600} of 0.8. Protein expression was induced for 16 h at 16 °C after adding 0.1 mM IPTG. Harvested cells were resuspended in lysis buffer (50 mM Tris pH 8, 200 mM NaCl) with protease inhibitor tablet (complete EDTA-free; Roche), lysozyme (Roche) and DNase I (Sigma-Aldrich). The cells were lysed by sonication and centrifuged at 10 000 ***g*** for 15 min. The membrane fraction was separated from soluble fraction after centrifugation at 200 000 ***g*** for 45 min. The membranes were solubilised in 50 mM Tris pH 8, 200 mM NaCl and 1% DDM at 22 °C for 30 min. The suspension was centrifuged at 200 000 ***g*** for 45 min. The supernatant was loaded onto a HisTrap column (GE Healthcare). The protein was eluted with a 0–100% gradient of buffer (50 mM Tris pH 8, 200 mM NaCl, 0.004% DDM, 500 mM imidazole). The His_6-MBP-tag of CagI was cleaved by the His_6-TEV protease during dialysis against 50 mM Tris pH 8, 200 mM NaCl and 0.004% DDM. The dialysed protein solution was loaded onto a HisTrap column and MBPTrap column (GE Healthcare). The cleaved proteins were collected in the flowthrough. The cleaved protein fractions were pooled, concentrated and loaded onto a Superdex 200 Increase 10/300 GL gel filtration column (GE Healthcare) equilibrated in 50 mM Tris pH 8, 200 mM NaCl and 0.0017% DMNPG.

CagF

The open reading frame encoding for CagF (*hp0543,* strain 26695) was inserted into pRSF-MBP. BL21 (DE3) cells

carrying the CagF expression vector were grown in LB at 37 °C until the OD_{600} reached 0.6 and at 37 °C until an OD_{600} of 0.8, and protein expression was induced for 16 h at 20 °C by adding 1 mM IPTG. Cells were harvested by centrifugation at 7000 ***g*** and resuspended in lysis buffer [50 mM Tris pH 8, 300 mM NaCl, 10% glycerol (v/v), 1% Triton (v/v)] with protease inhibitor tablet (complete EDTA-free; Roche), lysozyme (Roche) and DNase I (Sigma-Aldrich). The cells were lysed by sonication and centrifuged at 20 000 ***g*** for 15 min. The soluble fraction was loaded onto Ni-NTA resin (GE Healthcare). The resin was washed with 50 mL of buffer (50 mM Tris pH 8, 200 mM NaCl, 5% glycerol) and incubated overnight with 1 mg of TEV protease under agitation. The cleaved proteins was collected in the flowthrough fraction and subsequently purified on a Superdex 200 Increase 10/300 GL gel filtration column (GE Healthcare) equilibrated in 50 mM Tris pH 8, 200 mM NaCl and 5% glycerol.

CagY

The DNA sequence corresponding to $CagY^{B10}$ domain (residues 1542 to 1896, strain P12) was inserted into the pET51D vector (Invitrogen). BL21(DE3) cells carrying the expression vector were gown in LB with ampicillin (100 $g{\cdot}L^{-1}$), and expression was induced for 16 h at 20 °C by the addition of 1 mM IPTG. Purification was performed on a HisTrap column (GE Healthcare) using the same protocol than $CagA^{NTD}$. SEC on a Superdex S200 Increase 10/300 GL column (GE Healthcare) was performed in 20 mM Tris pH 8 and 200 mM NaCl.

CagY repeat II region ($CagY^{RRII}$, residues 464–1460, strain 26695) was inserted into the pGEX4T (GE Healthcare) vector in fusion with a N-terminal GST-tag cleavable by thrombin. BL21(DE3) cells harbouring the expression plasmid were grown into LB, and protein expression was induced for 16 h at 20 °C. After cell lysis in PBS and centrifugation at 20 000 ***g***, the supernatant was applied to a 5-mL GST column (Macherey-Nagel) and washed extensively with PBS. On-column GST-tag digestion was carried out during 4 h at 20 °C by injecting thrombin (150 units) in 20 mM Tris pH 8.4, 150 mM NaCl and 2.5 mM $CaCl_2$. The $CagY^{RRII}$ was eluted from the column using 20 mM Tris pH 7. The protein was then loaded onto a HiTrapSP HP column (GE Healthcare) and eluted using a linear gradient of buffer containing 20 mM Tris pH 7 and 1 M NaCl. SEC was performed on a Superdex 200 Increase 10/300 GL column in 20 mM Tris pH 7 and 100 mM NaCl.

Multi-angle laser 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 10/300 GL column (GE Healthcare) for $CagY^{RRII}$, $CagY^{B10}$ and CagL and on a Superdex S200 Increase 5/150 GL column for CagI. Experiments were performed with size-exclusion buffer of the corresponding proteins. Twenty-five microlitres of proteins was injected at a concentration of 10 $mg{\cdot}mL^{-1}$. 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.).

CD

Chirascan CD Spectrometer (Applied Photophysics) was used to record all Far-UV CD spectra (190–260 nm). Measurements were carried out at 20 °C in a 0.1-cm-pathlength quartz cuvette. Parameters were set as follows: wavelength range 190–260 nm, 0.2 nm increment, bandwidth 0.5 nm; scan speed, 50 $nm{\cdot}min^{-1}$; and response time, 1s. Spectra were corrected by subtracting buffer contributions and protein dilution factors before calculating the mean residue molar ellipticity. Spectra deconvolution was performed using DichroWeb server [48] and the Cdstr algorithm [49].

Small-angle X-ray scattering

Small-angle X-ray scattering data were recorded on SWING beamline at Synchrotron SOLEIL (Gif sur Yvette, France). Twenty-five microlitres of sample in buffer (50 mM Tris/HCl buffer pH 7.4, 137 mM NaCl, 2.7 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$) was injected into a Superdex S200 Increase 10/300 GL column using an Agilent© HPLC system cooled at 288 K and eluted directly into the SAXS flowthrough capillary cell at a flow rate of 0.2 $mL{\cdot}min^{-1}$. SAXS data were collected online throughout the whole elution time, with a frame duration of 2 s and a dead time between frames of 1 s. A first data set of 30 frames, collected before the void volume, was averaged to account for buffer scattering. A second data set was collected for the sample, from which the 10 frames corresponding to the top of the elution peak were averaged and used for data processing after baseline subtraction. Data were processed using the local application FOXTROT (http://www.synchrotron-soleil.fr/Recherche/LignesLumiere/SWING) and analysed using PRIMUS [50].

Surface plasmon resonance

Measurements were performed using a Biacore T200 instrument (GE Healthcare). Proteins 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^{-1} of a mixture of 200 mM *N*-ethyl-*N*′-(3-dimethylaminopropyl)carbodi-imide (EDC) and 50 mM *N*-hydroxysuccinimide (NHS). Proteins were diluted in 10 mM sodium acetate buffer at pH 4.0 for CagL, CagF, CagA^{SLB}, BSA and fibronectin; pH 4.5 for CagI and CagY^{B10}; and pH 5.5 for CagA^{NTD} and CagA before injection over the activated surface of the sensorchip. Residual active groups were blocked by injection of 1 M ethanolamine (pH 8.5). Immobilisation levels of 1000 RU (response units) were obtained for each protein. 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. Proteins were injected at 30 μL·min^{-1} for 200 seconds with a running buffer containing 50 mM Tris pH 7.4, 137 mM NaCl, 2.7 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$ and 0.05% P20. The sensorchip surface was regenerated with a pulse of 10 mM NaOH. The equilibrium K_D values were calculated using the BIAevaluation 3.0 software.

Acknowledgements

This project was funded by the ANR-13-ISV3-0006-Sintesys and DFG (HA2697/17-1) international programme, the Fondation pour la Recherche Médicale (FRM) and Agence Recherche contre le Cancer (ARC). We thank Roland Montserret for assistance with circular dichroism experiments. We acknowledge the use of the Protein Science Facility of the SFR Biosciences (UMS3444/US8). We thank staff from the SWING beamline at Synchrotron SOLEIL for help with SAXS data collection and processing.

Author contributions

TK, CB, MVC, KB, LT, RM and PR performed experiments; MVC, LB, LT, RH, WF and RF analysed the data; JT, LJS, RH and WF contributed reagents or other essential material; and TK, CB and LT wrote the manuscript.

References

1 Trokter M, Felisberto-Rodrigues C, Christie PJ & Waksman G (2014) Recent advances in the structural and molecular biology of type IV secretion systems. *Curr Opin Struct Biol* **27C**, 16–23.

2 Backert S, Tegtmeyer N & Fischer W (2015) Composition, structure and function of the *Helicobacter pylori* cag pathogenicity island encoded type IV secretion system. *Future Microbiol* **10**, 955–965.

3 Odenbreit S, Puls J, Sedlmaier B, Gerland E, Fischer W & Haas R (2000) Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* **287**, 1497–1500.

4 Mueller D, Tegtmeyer N, Brandt S, Yamaoka Y, De Poire E, Sgouras D, Wessler S, Torres J, Smolka A & Backert S (2012) c-Src and c-Abl kinases control hierarchic phosphorylation and function of the CagA effector protein in Western and East Asian *Helicobacter pylori* strains. *J Clin Invest* **122**, 1553–1566.

5 Hatakeyama M (2014) *Helicobacter pylori* CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. *Cell Host Microbe* **15**, 306–316.

6 Berge C & Terradot L (2017) Structural insights into *Helicobacter pylori* Cag protein interactions with host cell factors. *Curr Top Microbiol Immunol* **400**, 129–147.

7 Luo BH, Carman CV & Springer TA (2007) Structural basis of integrin regulation and signaling. *Annu Rev Immunol* **25**, 619–647.

8 Springer TA & Dustin ML (2012) Integrin inside-out signaling and the immunological synapse. *Curr Opin Cell Biol* **24**, 107–115.

9 Su Y, Xia W, Li J, Walz T, Humphries MJ, Vestweber D, Cabanas C, Lu C & Springer TA (2016) Relating conformation to function in integrin alpha5beta1. *Proc Natl Acad Sci USA* **113**, E3872–E3881.

10 Liu H, Lee MJ, Solis NV, Phan QT, Swidergall M, Ralph B, Ibrahim AS, Sheppard DC & Filler SG (2016) Aspergillus fumigatus CalA binds to integrin alpha5beta1 and mediates host cell invasion. *Nat Microbiol* **2**, 16211.

11 Hussein HA, Walker LR, Abdel-Raouf UM, Desouky SA, Montasser AK & Akula SM (2015) Beyond RGD: virus interactions with integrins. *Arch Virol* **160**, 2669–2681.

12 Hauck CR, Borisova M & Muenzner P (2012) Exploitation of integrin function by pathogenic microbes. *Curr Opin Cell Biol* **24**, 637–644.

13 Stewart PL & Nemerow GR (2007) Cell integrins: commonly used receptors for diverse viral pathogens. *Trends Microbiol* **15**, 500–507.

14 Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, Misselwitz R, Berger J, Sewald N, Konig W *et al.* (2007) Helicobacter exploits integrin for type IV secretion and kinase activation. *Nature* **449**, 862–866.

15 Schuelein R, Everingham P & Kwok T (2011) Integrin-mediated type IV secretion by Helicobacter: what makes it tick? *Trends Microbiol* **19**, 211–216.

16 Tegtmeyer N, Hartig R, Delahay RM, Rohde M, Brandt S, Conradi J, Takahashi S, Smolka AJ, Sewald N & Backert S (2010) A small fibronectin-mimicking protein from bacteria induces cell spreading and focal adhesion formation. *J Biol Chem* **285**, 23515–23526.

17 Saha A, Backert S, Hammond CE, Gooz M & Smolka AJ (2010) *Helicobacter pylori* CagL activates ADAM17

to induce repression of the gastric H, K-ATPase alpha subunit. *Gastroenterology* **139**, 239–248.
18 Wiedemann T, Hofbaur S, Tegtmeyer N, Huber S, Sewald N, Wessler S, Backert S & Rieder G (2012) *Helicobacter pylori* CagL dependent induction of gastrin expression via a novel alphavbeta5-integrin-integrin linked kinase signalling complex. *Gut* **61**, 986–996.
19 Conradi J, Tegtmeyer N, Wozna M, Wissbrock M, Michalek C, Gagell C, Cover TL, Frank R, Sewald N & Backert S (2012) An RGD helper sequence in CagL of *Helicobacter pylori* assists in interactions with integrins and injection of CagA. *Front Cell Infect Microbiol* **2**, 70.
20 Conradi J, Huber S, Gaus K, Mertink F, Royo Gracia S, Strijowski U, Backert S & Sewald N (2012) Cyclic RGD peptides interfere with binding of the *Helicobacter pylori* protein CagL to integrins alphaVbeta3 and alpha5beta1. *Amino Acids* **43**, 219–232.
21 Jimenez-Soto LF, Kutter S, Sewald X, Ertl C, Weiss E, Kapp U, Rohde M, Pirch T, Jung K, Retta SF *et al.* (2009) *Helicobacter pylori* type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner. *PLoS Pathog* **5**, e1000684.
22 Rohde M, Puls J, Buhrdorf R, Fischer W & Haas R (2003) A novel sheathed surface organelle of the *Helicobacter pylori* cag type IV secretion system. *Mol Microbiol* **49**, 219–234.
23 Kumar N, Shariq M, Kumari R, Tyagi RK & Mukhopadhyay G (2013) Cag type IV secretion system: CagI independent bacterial surface localization of CagA. *PLoS ONE* **8**, e74620.
24 Pham KT, Weiss E, Jimenez Soto LF, Breithaupt U, Haas R & Fischer W (2012) CagI is an essential component of the *Helicobacter pylori* Cag type IV secretion system and forms a complex with CagL. *PLoS ONE* **7**, e35341.
25 Shaffer CL, Gaddy JA, Loh JT, Johnson EM, Hill S, Hennig EE, McClain MS, McDonald WH & Cover TL (2011) *Helicobacter pylori* exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface. *PLoS Pathog* **7**, e1002237.
26 Kaplan-Turkoz B, Jimenez-Soto LF, Dian C, Ertl C, Remaut H, Louche A, Tosi T, Haas R & Terradot L (2012) Structural insights into *Helicobacter pylori* oncoprotein CagA interaction with beta1 integrin. *Proc Natl Acad Sci USA* **109**, 14640–14645.
27 Jimenez-Soto LF & Haas R (2016) The CagA toxin of *Helicobacter pylori*: abundant production but relatively low amount translocated. *Sci Rep* **6**, 23227.
28 Nagae M, Re S, Mihara E, Nogi T, Sugita Y & Takagi J (2012) Crystal structure of alpha5beta1 integrin ectodomain: atomic details of the fibronectin receptor. *J Cell Biol* **197**, 131–140.
29 Takagi J, Erickson HP & Springer TA (2001) C-terminal opening mimics 'inside-out' activation of integrin alpha5beta1. *Nat Struct Biol* **8**, 412–416.
30 Delahay RM, Balkwill GD, Bunting KA, Edwards W, Atherton JC & Searle MS (2008) The highly repetitive region of the *Helicobacter pylori* CagY protein comprises tandem arrays of an alpha-helical repeat module. *J Mol Biol* **377**, 956–971.
31 Chandran V, Fronzes R, Duquerroy S, Cronin N, Navaza J & Waksman G (2009) Structure of the outer membrane complex of a type IV secretion system. *Nature* **462**, 1011–1015.
32 Barden S, Lange S, Tegtmeyer N, Conradi J, Sewald N, Backert S & Niemann HH (2013) A helical RGD motif promoting cell adhesion: crystal structures of the *Helicobacter pylori* type IV secretion system pilus protein CagL. *Structure* **21**, 1931–1941.
33 Terradot L, Bayliss R, Oomen C, Leonard GA, Baron C & Waksman G (2005) Structures of two core subunits of the bacterial type IV secretion system, VirB8 from Brucella suis and ComB10 from *Helicobacter pylori*. *Proc Natl Acad Sci USA* **102**, 4596–4601.
34 Couturier MR, Tasca E, Montecucco C & Stein M (2006) Interaction with CagF is required for translocation of CagA into the host via the *Helicobacter pylori* type IV secretion system. *Infect Immun* **74**, 273–281.
35 Pattis I, Weiss E, Laugks R, Haas R & Fischer W (2007) The *Helicobacter pylori* CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. *Microbiology* **153**, 2896–2909.
36 Bonsor DA, Weiss E, Iosub-Amir A, Reingewertz TH, Chen TW, Haas R, Friedler A, Fischer W & Sundberg EJ (2013) Characterization of the translocation-competent complex between the *Helicobacter pylori* oncogenic protein CagA and the accessory protein CagF. *J Biol Chem* **288**, 32897–32909.
37 Bonig T, Olbermann P, Bats SH, Fischer W & Josenhans C (2016) Systematic site-directed mutagenesis of the *Helicobacter pylori* CagL protein of the Cag type IV secretion system identifies novel functional domains. *Sci Rep* **6**, 38101.
38 Bonsor DA, Pham KT, Beadenkopf R, Diederichs K, Haas R, Beckett D, Fischer W & Sundberg EJ (2015) 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* **290**, 12929–12940.
39 Barden S & Niemann HH (2015) Adhesion of several cell lines to *Helicobacter pylori* CagL is mediated by

integrin alphaVbeta6 via an RGDLXXL motif. *J Mol Biol* **427**, 1304–1315.

40 Rivera-Calzada A, Fronzes R, Savva CG, Chandran V, Lian PW, Laeremans T, Pardon E, Steyaert J, Remaut H, Waksman G *et al*. (2013) Structure of a bacterial type IV secretion core complex at subnanometre resolution. *EMBO J* **32**, 1195–1204.

41 Fronzes R, Schafer E, Wang L, Saibil HR, Orlova EV & Waksman G (2009) Structure of a type IV secretion system core complex. *Science* **323**, 266–268.

42 Terradot L & Waksman G (2011) Architecture of the *Helicobacter pylori* Cag-type IV secretion system. *FEBS J* **278**, 1213–1222.

43 Frick-Cheng AE, Pyburn TM, Voss BJ, McDonald WH, Ohi MD & Cover TL (2016) Molecular and structural analysis of the *Helicobacter Pylori* cag type IV secretion system core complex. *MBio* **7**, e02001-15.

44 Takagi J, Strokovich K, Springer TA & Walz T (2003) Structure of integrin alpha5beta1 in complex with fibronectin. *EMBO J* **22**, 4607–4615.

45 Busch B, Weimer R, Woischke C, Fischer W & Haas R (2015) *Helicobacter pylori* interferes with leukocyte migration via the outer membrane protein HopQ and via CagA translocation. *Int J Med Microbiol* **305**, 355–364.

46 Angelini A, Tosi T, Mas P, Acajjaoui S, Zanotti G, Terradot L & Hart DJ (2009) Expression of *Helicobacter pylori* CagA domains by library-based construct screening. *FEBS J* **276**, 816–824.

47 Hayashi T, Senda M, Morohashi H, Higashi H, Horio M, Kashiba Y, Nagase L, Sasaya D, Shimizu T, Venugopalan N *et al*. (2012) Tertiary structure-function analysis reveals the pathogenic signaling potentiation mechanism of *Helicobacter pylori* oncogenic effector CagA. *Cell Host Microbe* **12**, 20–33.

48 Whitmore L & Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, W668–W673.

49 Sreerama N, Venyaminov SY & Woody RW (1999) Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci* **8**, 370–380.

50 Konarev PV, Volkov VV, Sokolova AV, Koch MHJ & Svergun DI (2003) PRIMUS: a windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277–1282.