A Helical RGD Motif Promoting Cell Adhesion: Crystal Structures of the *Helicobacter pylori* Type IV Secretion System Pilus Protein CagL

Stephan Barden, Stefanie Lange, Nicole Tegtmeier, Jens Conradi, Norbert Sewald, Steffen Backert, and Hartmut H. Niemann

1 Structural Biochemistry, Department of Chemistry, Bielefeld University, 33501 Bielefeld, Germany
2 Organic and Bioorganic Chemistry, Department of Chemistry, Bielefeld University, 33501 Bielefeld, Germany
3 Institute of Medical Microbiology, OvG University Magdeburg, Leipziger Str. 44, D-39120 Magdeburg, Germany
4 Department Biology, FAU University Erlangen-Nuremberg, D-91058 Erlangen, Germany

**SUMMARY**

RGD tripeptide motifs frequently mediate ligand binding to integrins. The type IV secretion system (T4SS) protein CagL of the gastric pathogen *Helicobacter pylori* also contains an RGD motif. CagL decorates the T4SS pilus and may function as an adhesin for host cells. Whether CagL binds integrins via its RGD motif is under debate. Here, we present crystal structures of CagL revealing an elongated four-helix bundle that appears evolutionarily unrelated to the proposed VirB5 orthologs. The RGD motif is surface-exposed but located within a long α helix. This is unprecedented as previously characterized integrin-binding RGD motifs are located within extended or flexible loops. Yet, adhesion of gastric epithelial cells to CagL was strictly RGD-dependent. Comparison of seven crystallographically independent molecules reveals substantial structural flexibility. Intramolecular disulfide bonds engineered to reduce CagL flexibility resulted in more stable protein, but unable to support cell adhesion. CagL may thus partly unfold during receptor binding.

**INTRODUCTION**

Adhesion of animal cells to the extracellular matrix (ECM) is a fundamental process in development, tissue homeostasis, and disease. Specific cell surface receptors including the heterodimeric integrins mediate interactions with ECM proteins including fibronectin (Fn; Hynes, 2002; Singh et al., 2010). Fn alone is sufficient to induce spreading of various mammalian cell types. The Arg-Gly-Asp (RGD) tripeptide motif of Fn is crucial for integrin binding and RGD motifs are also present in other integrin ligands (Leiss et al., 2008; Ruslohtli, 1996). Some bacterial and viral pathogens target integrins, facilitating extracellular persistence or entry into host cells (Hauck et al., 2006; Stewart and Nemerow, 2007; Ulanova et al., 2009). One example is the gastric pathogen and group 1 carcinogen *Helicobacter pylori*, which exploits integrin receptors via a type IV secretion system (T4SS; Kwok et al., 2007). T4SSs are macromolecular assemblies that transport proteins or protein-DNA complexes across the bacterial envelope, resulting in secretion into the medium or transfer into bacterial or eukaryotic target cells (Alvarez-Martinez and Christie, 2009; Fronzes et al., 2009). T4SSs consist of three interlinked parts: a cytoplasmic/inner membrane-associated complex composed of three NTPases providing the energy for transport; a channel spanning the bacterial envelope, also known as the core complex; and an external pilus. The *Agrobacterium tumefaciens* T-DNA transfer system is the prototypic T4SS. It consists of 11 essential VirB proteins (VirB1-VirB11) and VirD4. The virulence-associated T4SS of *H. pylori* is encoded in an ~40 kb genomic insertion, the cytotoxin-associated gene pathogenicity island (cagPAI) (Odenbreit et al., 2000). Homologs for most of the 12 proteins of the *A. tumefaciens* VirB/D T4SS have been identified among the ~27 proteins encoded in the cagPAI, although some are poorly defined (Buhdorf et al., 2003; Covacci et al., 1997). Little is known about most remaining accessory proteins encoded by the cagPAI that lack homology to VirB/D proteins (Cendron and Zanotti, 2011).

CagL is one such cagPAI protein that lacks clear sequence similarity to any VirB/D protein. CagL localizes to the surface of the T4SS pilus (Kwok et al., 2007) and a ΔcagL mutant fails to translocate the oncogenic effector protein CagA and induce interleukin-8 (IL-8) secretion in host cells (Fischer et al., 2001). Despite a mere 15% sequence identity, CagL was suggested to be an ortholog of VirB5, the prototypical adhesin of the VirB/D system primarily based on its function, size (~220 amino acids), isoelectric point, genomic organization, and three-dimensional modeling (Backert et al., 2008). CagL was alternatively suggested to be a *H. pylori*-specific protein that shares sequence similarity with CagH (Shaffer et al., 2011). CagL can be copurified with CagH and CagI, two cagPAI proteins encoded within the same operon as CagL (Pham et al., 2012; Shaffer et al., 2011). All three share a conserved C-terminal “S/T-K-I/V-I-V-K”-hexapeptide and influence pilus formation (Shaffer et al., 2011).

CagL is the only cagPAI protein with an RGD motif and may function as a T4SS adhesin that binds to the host integrins α5β1, αvβ5, and αvβ3 (Conradi et al., 2012a; Jiménez-Soto et al., 2009; Kwok et al., 2007; Wiedemann et al., 2012),...
The structure of CagL (amino acids 21–237) was determined in two crystal forms. To reduce surface entropy and aid crystallization (Derewna, 2004), we replaced potentially surface exposed, high entropy amino acids by smaller, low entropy residues to generate the variant CagLKKQEK (Table 1). Functionality of CagLKKQEK in the T4SS was assessed by genetic complementation of CagL in H. pylori and has been proven to be as active as CagLwt (Figure S1 available online). The structure of CagLKKQEK was solved using seleno-methionine (SeMet) derivatized protein by multi-wavelength anomalous dispersion (MAD; Table 2). Cubic crystals contained one molecule per asymmetric unit. Following reductive methylation of primary amines of wild-type CagL (CagLmeth), we obtained a second, tetragonal crystal form with six molecules in the asymmetric unit. The structure was solved by molecular replacement (MR) using CagLKKQEK as search model. Data processing and refinement statistics for both crystal forms are provided in Table 3. While the six crystallographically independent molecules of CagLmeth are similar to each other (Table S1), there are major differences to CagLKKQEK. We will use the best defined molecule A of CagLmeth as a reference, and describe important differences for the higher resolution structure of CagLKKQEK where appropriate.

### CagL Forms an Elongated Helical Bundle

CagL is rod-shaped with dimensions of about 90 Å by 30 Å and consists of six α helices. Four long α helices (α1, α2, α5, and α6) run roughly parallel to the long axis with two short perpendicular helices (α3 and α4) winding around the N-terminal part of α5 (Figures 1A and S1). We will refer to CagL with the long axis oriented vertically such that both N and C termini point up, positioning the short helices α3 and α4 at the top. The connections between α1 and α2 and between α5 and α6 form the bottom of the molecule. Viewed from the top, α1, α2, α5, and α6 are arranged counterclockwise. The elongated shape and mainly α helical structure of CagL are consistent with the molecular mass estimate of ~50 kDa from analytical gel filtration (calculated mass from sequence ~25 kDa) and with circular dichroism (CD) spectra of CagL, respectively (Figures S2A and S3A; Conradi et al., 2012b; Kwok et al., 2007).

Although the fold is clearly defined, CagL is characterized by substantial structural flexibility. The most invariant part of CagL is a three-helix bundle formed by the upper half of α2, α5, and α6 together with α3 (Figures 1A and 1B). This CagL core is held together by a cluster of aromatic residues: Phe86, Phe92, Phe93, and Phe100 from α2, Tyr163 from α5, and His204, Phe204, and Tyr207 from α6. Minor structural variations are found in the C terminus of α5, the N terminus of α6, and the connecting short loop. This slightly more flexible region begins with a double glycine (Gly168, Gly169) in α5 and ends with Leu184 in α6. The N-terminal part of α2 extends over the loop connecting α5 and α6 and does not contribute to the three-helix bundle. Comparison of the seven independent molecules reveals a hinge region in the central part of α2 around Gly77 of the RGD motif, which allows the N-terminal part of α2 to move toward the C terminus of α5 (Figure 1B). This structural flexibility occurs between the two different crystal forms and also between different

---

### RESULTS

**Crystallization and Structure Determination of CagL**

The structure of CagL (amino acids 21–237) was determined in two crystal forms. To reduce surface entropy and aid crystallization (Derewna, 2004), we replaced potentially surface exposed, high entropy amino acids by smaller, low entropy residues to generate the variant CagLKKQEK (Table 1). Functionality of CagLKKQEK in the T4SS was assessed by genetic complementation of CagL in H. pylori and has been proven to be as active as CagLwt (Figure S1 available online). The structure of CagLKKQEK was solved using seleno-methionine (SeMet) derivatized protein by multi-wavelength anomalous dispersion (MAD; Table 2). Cubic crystals contained one molecule per asymmetric unit. Following reductive methylation of primary amines of wild-type CagL (CagLmeth), we obtained a second, tetragonal crystal form with six molecules in the asymmetric unit. The structure was solved by molecular replacement (MR) using CagLKKQEK as search model. Data processing and refinement statistics for both crystal forms are provided in Table 3. While the six crystallographically independent molecules of CagLmeth are similar to each other (Table S1), there are major differences to CagLKKQEK. We will use the best defined molecule A of CagLmeth as a reference, and describe important differences for the higher resolution structure of CagLKKQEK where appropriate.

CagL Forms an Elongated Helical Bundle

CagL is rod-shaped with dimensions of about 90 Å by 30 Å and consists of six α helices. Four long α helices (α1, α2, α5, and α6) run roughly parallel to the long axis with two short perpendicular helices (α3 and α4) winding around the N-terminal part of α5 (Figures 1A and 1B). We will refer to CagL with the long axis oriented vertically such that both N and C termini point up, positioning the short helices α3 and α4 at the top. The connections between α1 and α2 and between α5 and α6 form the bottom of the molecule. Viewed from the top, α1, α2, α5, and α6 are arranged counterclockwise. The elongated shape and mainly α helical structure of CagL are consistent with the molecular mass estimate of ~50 kDa from analytical gel filtration (calculated mass from sequence ~25 kDa) and with circular dichroism (CD) spectra of CagL, respectively (Figures S2A and S3A; Conradi et al., 2012b; Kwok et al., 2007).

Although the fold is clearly defined, CagL is characterized by substantial structural flexibility. The most invariant part of CagL is a three-helix bundle formed by the upper half of α2, α5, and α6 together with α3 (Figures 1A and 1B). This CagL core is held together by a cluster of aromatic residues: Phe86, Phe92, Phe93, and Phe100 from α2, Tyr163 from α5, and His204, Phe204, and Tyr207 from α6. Minor structural variations are found in the C terminus of α5, the N terminus of α6, and the connecting short loop. This slightly more flexible region begins with a double glycine (Gly168, Gly169) in α5 and ends with Leu184 in α6. The N-terminal part of α2 extends over the loop connecting α5 and α6 and does not contribute to the three-helix bundle. Comparison of the seven independent molecules reveals a hinge region in the central part of α2 around Gly77 of the RGD motif, which allows the N-terminal part of α2 to move toward the C terminus of α5 (Figure 1B). This structural flexibility occurs between the two different crystal forms and also between different
Structure of HP CagL Reveals a Helical RGD Motif

### Table 2. Data Collection and Phasing Statistics

<table>
<thead>
<tr>
<th>Protein</th>
<th>CagL&lt;sub&gt;KKQEK&lt;/sub&gt; (SeMet)</th>
<th>CagL&lt;sub&gt;meth&lt;/sub&gt; (Native)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>BL14.2 (BESSY)</td>
<td>ID23.2 (ESRF)</td>
</tr>
<tr>
<td>Space group</td>
<td>F23</td>
<td>F23</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97814/0.978247</td>
<td>0.8726</td>
</tr>
<tr>
<td>Unit cell axis (Å)</td>
<td>a = b = c = 181.24, 181.22</td>
<td>a = b = c = 181.24, 181.22</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20–2.2 (2.29–2.2)</td>
<td>20–2.2 (2.29–2.2)</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>24,942 (3,606)</td>
<td>260,650 (37,776)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Anomalous multiplicity</td>
<td>14.4 (9.2)</td>
<td>7.7 (7.8)</td>
</tr>
<tr>
<td>Anomalous completeness (%)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.069 (0.234)</td>
<td>0.139 (0.908)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.076 (0.257)</td>
<td>0.144 (0.962)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.031 (0.106)</td>
<td>0.037 (0.317)</td>
</tr>
<tr>
<td>I/σ</td>
<td>22.2 (9.3)</td>
<td>12.7 (2.3)</td>
</tr>
<tr>
<td>Mosaicity (%)</td>
<td>0.54</td>
<td>0.73</td>
</tr>
<tr>
<td>delAnom between half sets</td>
<td>0.700 (0.213)</td>
<td>0.998 (0.716)</td>
</tr>
<tr>
<td>Phasing (after three cycles bucaneer/refmac)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation coefficient all/weak (%)</td>
<td>55.04/40.05</td>
<td></td>
</tr>
<tr>
<td>Site occupancy #1/#6 (search for five sites)</td>
<td>1.00/0.20</td>
<td></td>
</tr>
<tr>
<td>Contrast (after 20 cycles SHELXE)</td>
<td>1.23</td>
<td>23.623 (2,614)</td>
</tr>
<tr>
<td>Pseudo-free CC (%)</td>
<td>64.84</td>
<td>31,957 (2632)</td>
</tr>
<tr>
<td>Automated model building</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. chains</td>
<td>5 (longest 85 residues)</td>
<td></td>
</tr>
<tr>
<td>No. residues</td>
<td>190 (157 sequenced)</td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;cryst&lt;/sub&gt; (%)</td>
<td>31.86</td>
<td>0.012</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>34.35</td>
<td>1.29</td>
</tr>
<tr>
<td>Wilson B-factor (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>46.2</td>
<td>91.1</td>
</tr>
</tbody>
</table>

### Helix α1 Packs Loosely against the Three-Helix Bundle

The two crystal structures of CagL differ with respect to binding of the N-terminal helix to the three-helix bundle. α1 packs into a hydrophobic groove formed by ζ2 and ζ6 with minor contributions from ζ5 (Leu173 and Ala176) in both structures. In molecules of CagL<sub>meth</sub>, α3 is held in place by π−π-stacking of the phenolic rings of Tyr103 (C terminus of α2) and Tyr113 (α3) and a hydrogen bond between the amide nitrogen of Leu119 (C terminus of α3) and the side chain of Glu140 (α5; Figure 1C). α4 is covalently bound to the three-helix bundle by a disulfide bond between Cys128 from α4 and Cys139 from α5. Additionally, a hydrogen bond between Asp132 from the very top of α5 and Tyr225 from α6 stabilizes the structural arrangement in this region. Nevertheless, α4 remains quite flexible and was visible in only three of the seven crystallographically independent molecules, indicating the possibility of partial reduction of the disulfide bond.

CagL<sub>meth</sub>, α1 interacts with the more rigid upper part of the three-helix bundle. In CagL<sub>KKQEK</sub>, by contrast, α1 is shifted one helix turn toward the N terminus of α2 and α6 (Figure 2). In both structures, amino acids connecting α1 and α2 were disordered. The distance between the last ordered residue of α1 and the first ordered residue of α2 is ~17 Å in CagL<sub>meth</sub> (from Thr51 to Asn58) and CagL<sub>KKQEK</sub> (from Ser50 to Glu59). The missing residues are sufficient to span this distance in both cases.

In vitro, CagL was cleaved over time into a larger C-terminal and a smaller N-terminal fragment containing α1 (Figure S2). In gel filtration, the two fragments eluted together, suggesting that they remain associated. A variant of CagL lacking the...
N-terminal helix (CagL\textsuperscript{DN}) precipitated at concentrations above \(\sim 1\) mg/ml whereas CagL\textsuperscript{wt} remained soluble even beyond 10 mg/ml in PBS at 4°C. \(\alpha_1\) presumably shields the hydrophobic core from solvent, preventing aggregation. Consistently, the buried surface area between \(\alpha_1\) and the rest of CagL measuring 1,555 Å\(^2\) for CagL\textsuperscript{meth} chain A and 1,686 Å\(^2\) for CagL\textsuperscript{KKQEK} is calculated to be \(\sim 82\%\) hydrophobic in both cases (Hubbard and Thornton, 1993). Moreover, the melting temperature of CagL\textsuperscript{DN} in differential scanning fluorimetry (DSF) experiments is lower than for CagL\textsuperscript{wt}, as is its ability to promote cell adhesion (see below). Despite its variable association with the rest of CagL, \(\alpha_1\) may thus be important for stability.

It is not clear whether one of the two observed positions of \(\alpha_1\) represents the physiological structure. We propose that the binding mode of CagL\textsuperscript{meth} represents the native state because (1) the same binding mode is seen in six crystallographically independent molecules; (2) in five of the six molecules of CagL\textsuperscript{meth} \(\alpha_1\) is not involved in crystal contacts at all, whereas the C terminus of \(\alpha_1\) interacts with a symmetrical molecule in CagL\textsuperscript{KKQEK}; and (3) in CagL\textsuperscript{KKQEK} two dioxane molecules added during

---

**Figure 1. CagL Forms an Elongated Helical Bundle**

(A) Left: cartoon representation of chain A from CagL\textsuperscript{meth} colored in blue to red from N to C termini. Amino acids connecting \(\alpha_1\) and \(\alpha_2\) (Thr52–Asn57) are not resolved and are denoted as a dashed line. The conformation of the N-terminal end of \(\alpha_2\) in CagL\textsuperscript{KKQEK} is superimposed in transparent coloring (alignment of aa 80–220). The helical Arg76–Gly77–Asp78 (RGD) motif is highlighted as magenta spheres. Right: the same molecule rotated by 55° counterclockwise around the long axis colored by B-factor from blue (low) to red (high). Aromatic amino acid side chains are depicted as sticks (see also Figures S1–S3).

(B) Superposition of the seven crystallographically independent molecules of CagL solved in this study. Least-squares fit superposition was done for main chain atoms of amino acids 80–114, 144–166, and 187–219 (see also Table S1).

(C) Structural features that stabilize the upper part of CagL. A disulfide bond links Cys128 to Cys139. Coloring as in (A) left.
facilitates formation of a crystal contact around a 2-fold symmetry axis between the Cβ atoms of Ala69 (Figure 3D). This crystal contact involves the Arg76 side chain, which contacts the C terminus of α1 of the symmetry-related molecule (Figures 3C and 3D). Distortion of the helical conformation around the RGD motif may additionally be linked to the shift of α1 in CagLKKQEK. This shift comes along with an H-bond between the side chain of Tyr34 (α1) and the carbonyl oxygen of Leu75 (Figure 3C), which does not form the typical α helical intra-main chain H-bond.

Adhesion of MKN-45 Cells to CagL Is Strictly RGD-Dependent
Because CagL is a ligand for integrins, we analyzed adhesion of human gastric adenocarcinoma (MKN-45) cells to plates coated with wild-type CagL or CagL variants, in which each of the three amino acids from the RGD motif was substituted separately (Kwok et al., 2007; Schneider et al., 2011; Tegtmeyer et al., 2010; Table 1). In wells coated with Fn or CagLwt, cells evenly distributed over the surface. In control wells or wells coated with any of the CagL RGD variants, cells formed aggregates in solution (Figure 4A) and were lost upon washing. Quantification by a hexosaminidase assay revealed that MKN-45 cells adhered to wells coated with 125–500 nM CagLwt as well as the crystallization variants CagLKKQEK and CagLmeth. Signal intensity similar to that of 25 nM Fn was observed for 500 nM CagL. No adhesion was observed to any of the RGD variants (Figure 4B).

Mutation of Residues around the RGD Motif Does Not Impair Cell Adhesion
The protein surface around the RGD motif is characterized by hydrophobic or small amino acids: Ile73 and Ala74 one turn N-terminal of the RGD motif as well as Ala80 or Ala81, Ala84 or Ala88 one, two, or three turns toward the C terminus (Figure 4C). This hydrophobic patch flanking the RGD motif could be critical for integrin binding. Residues on α5 adjacent to the RGD motif are small as well. Exposed amino acids include Ser165, Gly168, Gly169, Ala171, and Ser172. To assess the influence of the exposed amino acids in proximity to the RGD motif, CagL variants CagL173N, CagLAL80GA, CagLIE87AS, CagLGL169K, and CagLAS171SN (Table 1) were tested in cell adhesion assays (Figure 4D). All variants were active in promoting cell adhesion with a similar concentration profile as CagLwt. The C terminus of CagL was similarly tested for adhesion of MKN-45 cells (Figures 1C and 4D). Both CagLGL169K and CagLGAS171SN, resembling a reduced disulfide bond, and the C-terminal deletion variant CagLΔC, lacking the hexapeptide conserved in CagL, CagL, and CagH, retained full activity. In contrast, CagLY225A showed reduced adhesion, possibly caused by the reduced thermostability of CagLY225A due to the loss of the intramolecular interaction with Asp132. Finally, CagLCHIS, described previously (Kwok et al., 2007; Saha et al., 2010; Tegtmeyer et al., 2010; Wiedemann et al., 2012), supported equal adherence to CagLwt.

The N Terminus of α2 Lacks Helical Stability
Susceptibility to proteases can identify flexible regions. Trypsin digestion of CagLCHIS rapidly generated a slightly smaller fragment that no longer reacted with an antibody directed against...
the His-tag. Consistent with this C-terminal truncation, amino acids Ser232 to Lys237 were not visible in any of the seven molecules, suggesting that the C terminus of a6 including the conserved hexapeptide is flexible. Continued proteolysis of CagLC-HIS or CagLwt as well as long-term storage of CagL in the absence of exogenously added proteases resulted in formation of a dominant fragment of 18.3 kDa (Figures 5A and S2).

N-terminal sequencing suggested a cleavage C-terminal of Lys70 but was ambiguous. To distinguish between cleavage C-terminal of Lys69-Lys70 or after Arg76, we performed tryptic digests of CagLR76A and CagLKKQEK (containing the K69A and K70A mutations). CagLR76A yielded similar degradation products as CagLwt, while the 18.3 kDa fragment did not form with CagLKKQEK, indicating a cleavage C-terminal of Lys70 rather than Arg76 (Figure 5A). Because Lys70 is part of an α helix in all our structures, cleavage at this site was unexpected and suggests that the N-terminal section of α2 may be flexible and lack helical stability in solution.

Restraining Flexibility around the RGD Motif Abrogates Cell Adhesion
To investigate whether the flexible, protease sensitive part of α2 affects CagL stability, we introduced disulfide bonds (Hazes and Dijkstra, 1988) to link flexible α2 and α5. In cysteine variant CagLKKG, a cysteine located six amino acids N-terminal of Gly77 links A71C to S172C (Table 1; Figure 5B). In CagLC2 residue N85C, eight amino acids C-terminal of Gly77 is linked to A162C (Table 1; Figure 5B). Tryptic proteolysis of CagLC2 (central disulfide) was indistinguishable from that of CagLwt whereas CagLC1 (more restrained α2) was more resistant to trypsin resulting in a CagLKKQEK-like pattern. DSF (Figure 5C) and temperature interval-dependent CD-spectroscopic measurements (Figure S3B) demonstrate that the disulfide bond stabilizes CagLC1. It undergoes a structural transition at ~49°C whereas CagLwt melted at ~43°C. Interestingly, CagLR76A and CagLS77A also exhibited a slightly elevated transition temperature of ~45°C. The melting curve of CagLC2 was qualitatively different from that of CagLwt, showing a biphasic behavior, with the second transition and the maximum shifted toward higher temperature. This possibly reflects a stabilization of the hydrophobic core in proximity of the C2 mutation. A variant harboring both disulfides (CagLKKGK) combined melting characteristics of variants CagLC1 and CagLC2, indicating independent effects on CagL stability.

To assess an influence of the flexibility of α2 on the accessibility of the RGD motif, we performed cell adhesion assays...
with the disulfide-linked variants. Interestingly, the C1 variant was strongly attenuated in its ability to induce adhesion of MKN-45 cells and the C2 variant was completely inactive (Figure 5D), indicating that not only accessibility of the RGD motif is relevant for receptor binding, but also conformational flexibility of α2 around the RGD motif.

**DISCUSSION**

We determined the structure of CagL variants because our efforts to determine the structure of wild-type CagL had failed. Surface entropy reduction mutations and lysine methylation are two standard methods to increase the crystallization propensity of proteins or to improve crystal quality (Derewenda, 2004; Walter et al., 2006). Both CagLKKQEK and CagLmeth supported cell adhesion similar to CagLwt. CagLKKQEK also complemented a cagL deletion in *H. pylori*. Genetic complementation is not possible for CagLmeth, but generally, structures of methylated proteins are very similar to their nonmethylated counterparts (Sledz et al., 2010). Together, these results strongly suggest that CagLKKQEK and CagLmeth are physiologically relevant and that their structures provide biologically meaningful insight into the function of wild-type CagL.

In previous studies, CagL was proposed to be a VirB5 ortholog and a structural model of CagL was generated based on the crystal structure of TraC, a VirB5 protein (Backert et al., 2008). However, the structural data presented here clearly distinguish CagL from TraC (Yeo et al., 2003). First, CagL consists of an elongated four-helix bundle, whereas TraC contains a three-helix bundle with a loose globular appendage. Moreover, the arrangement of helices in CagL and TraC is inverted. Second, structural features such as aromatic residues of the hydrophobic core and the disulfide bond of CagL are not conserved in TraC. A DALI search (Holm and Rosenström, 2010) revealed TraC as largely unrelated with a Z-score of 4.9 (root-mean-square deviation [rmsd] of 10.2 Å over 130 aligned residues). More related structures include diverse helical bundles e.g., of STATs, spectrin repeats, α-actinin, or the large GTPase Mx1 and also the type III secretion translocator protein SipB from *Shigella flexneri* and hemolysin BL from *Bacillus cereus*. No obvious sequence

---

**Figure 4. Adhesion of MKN-45 Cells to Immobilized CagL Depends on the RGD Motif**

(A) Phase contrast microscopy image showing adhesion of MKN-45 cells to precoated microtiter plates after 2 hr incubation without washing. Five hundred nanomolar CagLwt (upper right; induced a similar microscopic appearance as 25 nM fn; upper left) with an even distribution of adhered cells on the well surface. In contrast, cells did not adhere to negative control wells (lower left) or CagLR76A coated wells (lower right), but aggregated in the medium. Bars represent 50 μm.

(B and D) Quantification of the adherence by a hexosaminidase enzyme assay is shown. Absorbances at 405 nm were normalized to the highest signal intensity at 25 nM Fn. Data presented are mean absorbances for four or more independent experiments with four replicates each and corresponding standard deviations (n ≥ 4). One-way ANOVA statistics (OriginPro Software, OriginLabs) compared to CagLwt: n.s., not significant; *p < 0.05; **p < 0.01; ***p < 0.001. (C) Localization of the structure-derived mutations for adhesion assays. Ca2+ atoms of selected amino acids surrounding the RGD motif are shown as spheres, side chains as sticks. Mutated amino acids are highlighted in magenta.
homology was found in any structure. Although CagL functionally resembles VirB5 proteins in other T4SSs, it is probably a protein specific for the H. pylori cagPAI, as has been previously suggested (Kutter et al., 2008; Shaffer et al., 2011).

The manner in which H. pylori exploits integrins for cell attachment and type IV secretion is still under discussion. An RGD-dependent model proposes CagL to be located on the surface of the T4SS pilus mediating the first contact with host cell integrins (Kwok et al., 2007; Tegtmeyer et al., 2010). A competing RGD-independent model proposes a direct interaction of CagA with integrin (Jiménez-Soto et al., 2009; Kaplan-Türköz et al., 2012). The data presented in this study confirm that CagL alone is sufficient to induce adhesion of human gastric epithelial MKN-45 cells. Moreover, cell adhesion was strictly dependent on the RGD motif because alanine substitutions of any one of the three amino acids abrogated cell adhesion whereas mutation of surface-exposed amino acids in proximity to the RGD motif did not result in loss of function.

RGD motifs are mostly located in exposed, flexible loop structures extending beyond the protein surface. The RGD motif of the tenth type III repeat of Fn e.g., is located in a β-hairpin-like loop extending ~10 Å beyond the core of the domain (Leahy et al., 1996). The RGD motif of protein developmental endothelial cell locus-1 (Del-1) forms the tip of a long protruding loop important for interaction with integrin (Schürpf et al., 2012). Functional RGD motifs also occur in flexible loops, as is the case in latent transforming growth factor-β, where part of the RGD motif is disordered (Shi et al., 2011). RGD motifs exposed by viral or bacterial proteins that mediate integrin binding in the infection cycle are likewise located in loop regions and are mostly flexible (Emsley et al., 1996; Kagawa et al., 2000; Logan et al., 1993; Zubieta et al., 2005). Clearly, the RGD motif of CagL does not mimic the loop location of other structurally described RGD motifs. Instead, it is embedded in a distinct structural context in the middle of a long helix. This raises the question of how integrins recognize CagL.

Conformational flexibility around the RGD motif of CagL proved to be important for CagL recognition by cells. Disulfide bonds introduced N-terminal or C-terminal of the RGD motif serve to stabilize the protein but abolish cell adhesion. This agrees with our finding that the N-terminal section of α2 lacks helical stability. However, it was surprising that the C2 variant did not support cell adhesion because this disulfide bridge is located near the hydrophobic core. Though it would appear unlikely that the hydrophobic core of CagL could unfold upon receptor binding, the effect of the C2 variant does suggest that this may be the case.
Insoluble residual was pelleted at 20,000 × g, cooled centrifuge, washed twice with PBS, and dissolved in PBS + 8 M urea. 25 mM imidazole. Refolding of CagL was performed at 4 °C successively with 1 M NaCl, 12.5% glycerol, and 10 mM dithiothreitol (DTT) or containing 1 g/l NH4Cl, 3 g/l KH2PO4, and 6 g/l Na2HPO4·7H2O supplemented CagL aa 21–237 (CagLwt) with a linker sequence between the His-tag and CagL sensitive to the tobacco etch virus (TEV) protease. The surface entropy reduction variant CagLKKQEK was identified by the SERRp server (Goldschmidt et al., 2007) and generated by a modified QuickChange protocol (Stratagene) using Phusion polymerase. Similar approaches were used for generation of the variants listed in Table 1. Truncations of the N or the C termini were generated by amplification of the corresponding vector fragment, followed by DpnI digestion and ligation with T4 DNA ligase. All constructs were verified by sequencing.

Protein Expression and Purification
All CagL variants were expressed in E. coli BL21-CodonPlus-RIL for 3.5 hr at 37°C in selective lysogeny broth (LB) supplemented with 2 g/l glucose monohydrate and purified from inclusion bodies. The expression of CagL was induced with 0.5 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an optical density 600 (OD600) = 0.8. Inclusion bodies were isolated at 5,000 × g in a cooled centrifuge, washed twice with PBS, and dissolved in PBS + 8 M urea. Insoluble residual was pelleted at 20,000 × g. Unfolded CagL was bound to a Ni-NTA affinity column and washed with PBS + 8 M urea supplemented successively with 1 M NaCl, 12.5% glycerol, and 10 mM dithiotreitol (DTT) or 25 mM imidazole. Refolding of CagL was performed at 4 °C with a 1:5-fold dilution series of PBS + 6 M urea with refolding buffer, containing 50 mM Tris(hydroxymethyl)-ammonomethane (Tris), 20 mM NaCl, 0.8 mM KCl, 2 mM reduced glutathione, and 0.2 mM oxidized glutathione at a final pH 8.2 at 4 °C. Five milliliters of each dilution was applied for seven steps, followed by 25 ml of refolding buffer. CagL was eluted with PBS + 250 mM imidazole. The fusion protein was cleaved with TEV protease during dialysis against PBS overnight at 4 °C. Cleaved CagL was further isolated by negative affinity chromatography using Ni-NTA. Protein used for cell adhesion assays was not cleaved. The flow-through or the dialyzed tagged CagL was concentrated by ultra-filtration (VIVASPIN 20, 5,000 MWCO, Sartorius) and applied to a Hiload 16/60 Superdex75 pg size exclusion chromatography column (GE Healthcare) using Tris-buffered saline (TBS) for protein intended for cell adhesion assays and ligation with T4 DNA ligase. All constructs were verified by sequencing.

Crystal Harvesting, Cryo-Protection, and Data Collection
Crystals of CagLKKQEK were sensitive to evaporation of liquid or 1,4-dioxane from the crystallization drop. A cryoprotection solution equivalent to the crystallization drop contents plus 20% (v/v) 2.3-butanediol was added to the crystallization drop and noncracked crystals were transferred to higher 2.3-butanediol concentration, mounted in a cryo loop (Hampton Research), and flash-cooled in liquid nitrogen. The crystallization drop of CagLmeth was mixed with glycerol to a final concentration of 20% (v/v) prior to flash-cooling. CagLmeth diffraction data were collected on BL14.2 operated by the Joint Berlin MX-Laboratory at the BESSY II electron storage ring (Berlin-Adlershof; Mueller et al., 2012). Data sets from CagLmeth were collected at beamline ID23-2 of the European Synchrotron Radiation Facility (ESRF).

Data Reduction, Structure Determination, and Refinement
The data sets were either processed with XDS (Kabsch, 2010) or iMosFLM 1.06 (Battye et al., 2011) and scaled with Scala (Evans, 2008) from the CCP4 suite (CCP4, 1994). SHELXC, -D, and -E (Sheldrick, 2010) via the HKL2MAP interface (Pape and Schneider, 2004) were used for anomalous scatterer localization and phasing of the SeMet derivative of CagLmeth. An initial model was built with Buccaneer (Cowtan, 2006). The structure of CagLmeth was solved by MR using Phaser (McCoy et al., 2007). The models were manually modified in COOT (Emsley et al., 2010) and refined with PHENIX (Adams et al., 2010). For CagLmeth, the refinement included secondary structure restraints, NCS restraints, and reference structure restraints. Translation/Libration/Screw (TLS) groups were identified with the TLSMD server (Painter and Merritt, 2006) and included in refinement. LSQKAB was used for structural alignments (Kabsch, 1978). Figures were prepared with the PyMOL Molecular Graphics System, Version 0.99rc8 (Schrodinger).

Cell Adhesion Assays
MKN-45 cells (Deutsche Sammlung von Mikroorganismen und Zellkultur, DSMZ, Braunschweig) were cultivated at 37°C with 5% (v/v) CO2 in RPMI 1640 medium (PAA) supplemented with 10% fetal calf serum (Lonza), 4 mM L-glutamine, 200 U/ml penicillin, and 200 μg/ml streptomycin. The medium was exchanged every 2–3 days recovering nonadherent cells for further cultivation by centrifugation.

For cell adhesion assays, each well of a transparent P96 MicroWell plate with MaxiSorp surface (Nunc) was coated overnight at 4°C with 50 μl 62.5–500 nM CagL or 2.5–25 nM recombinant human (rh) Fn (R&D Systems) diluted in PBS. The wells were then supplemented with 100 μl of 5% (v/v) fresh milk powder solution for 2 hr at 4°C and finally blocked with 200 μl of fresh milk powder solution for another 2 hr. Harvested nonattached cells and trypsinized detached cells were diluted to 4–5 × 105 cells/ml in fresh medium and 10 μl of each MgCl2/CaCl2/MnCl2 was added. The milk powder solution was discarded and 100 μl of the cell solution was applied. The plates were incubated without a lid for 2 hr at 37°C with 5% (v/v) CO2. For quantification of the adhered cells, a hexosaminidase enzyme assay (Landegren, 1984) was performed. Nonattached cells were removed by twice washing each well with 200 μl PBS. Sixty microliters of a solution containing 1.9 mM p-nitrophenol-N-acetyl-β-D-glucosaminide (Sigma-Aldrich), 0.25% (v/v) Triton X-100, and 50 mM citrate buffer at pH 5 was added and incubated at room temperature for 75 min. The reaction was stopped by adding 90 μl of 200 mM glycine-NaOH buffer pH 10.4. Resulting absorbances were measured in a Tecan Infinite200 microplate reader at 405 nm. Each plate contained 25 nM Fn as reference and all absorbance readings were normalized to this value after background subtraction.

Genetic Complementation of CagL H. pylori
Mutagenesis and genetic complementation of CagL in H. pylori strain P12, culturing and infection of AGS (gastric adenocarcinoma) cells, quantification of the elongation phenotype, and western blotting using anti-phospho-
CagA, anti-CagA and anti-HA antibodies was performed as described elsewhere (Kwok et al., 2007; Conradi et al., 2012b).

**Differential Scanning Fluorimetry**

DSF was measured using a StepOnePlus Real-Time PCR System (Applied Biosystems). Experiments were performed with TBS as dilution buffer in G060/H/1E 96 well PCR plates (Kisker Biotech) sealed with G480-OS optical film (Kisker Biotech). Protein (42.5 μl) at 0.0625, 0.125, and 0.25 mg/ml were mixed with 2.5 μl of 38X SYPRO Orange (Life Technologies) directly in the PCR plates on ice. The plates were centrifuged at 4°C for 5 min at 2,000 x g and transferred to the real-time PCR cycler. Fluorescence was measured using the ROX dye settings from 4°C to 97°C with a heating speed of 1°C/min. Curves were analyzed using the transformation and analysis files provided by Niesen (Niesen et al., 2007) and GraphPad Prism 4 (GraphPad Prism). All measurements were repeated at least three times with three wells per data point.

**CD Spectroscopy**

CD spectroscopy (Johnson, 1990) was performed in 1 mm path-length quartz cuvettes from 270 nm to 190 nm wavelength with a bandwidth of 1 nm at a 50 nm/min scanning speed using a J-810 spectropolarimeter (Jasco Instruments). Temperature interval-dependent CD spectra were acquired at 0.15 mg/ml protein in 1:4 diluted crystallization buffer from 25°C to 65°C (75°C) with one data point per each 2°C. A corresponding baseline was subtracted.

**ACCESSION NUMBERS**

The Protein Data Bank accession numbers for the model coordinates and structure factors of CagL<sub>QEQE</sub> and CagL<sub>meth</sub> are 3ZCI and 3ZCJ, respectively.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes four figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.08.018.

**AUTHOR CONTRIBUTIONS**

S. Barden and H.N. designed the research, analyzed the data, and wrote the paper; S. Barden and S.L. performed research; N.T. and S. Backert performed and evaluated complementation experiments with H. pylori; S. Backert, J.C., and N.S. initiated the collaboration and provided reagents.

**ACKNOWLEDGMENTS**

We thank Christina Geerds and Inge lore Brandes for excellent technical assistance. We thank the beamline staff at BL14.2 of the BESSY II electron storage ring (Berlin, Germany) and acknowledge coverage of travel expenses for data collection by the Helmholtz Zentrum Berlin. We gratefully acknowledge access to beamline ID23-2 of the ESRF (Grenoble, France) and coverage of travel expenses. The work of S. Backert was supported through two DFG grants (Ba1671/8-1 and project B10 of CRC-798).

Received: April 15, 2013
Revised: August 8, 2013
Accepted: August 17, 2013
Published: September 26, 2013

**REFERENCES**


Structure

Structure of HP CagL Reveals a Helical RGD Motif


