Cell Host & Microbe
Article

Tertiary Structure-Function Analysis Reveals the Pathogenic Signaling Potentiation Mechanism of *Helicobacter pylori* Oncogenic Effector CagA

Takeru Hayashi,[1,2] Miki Senda,[3] Hiroko Morohashi,[1] Hideaki Higashi,[4] Masafumi Horio,[4] Yui Kashiba,[1] Lisa Nagase,[1] Daisuke Sasaya,[4] Tomohiro Shimizu,[4] Nagarajan Venugopalan,[5] Hiroyuki Kumeta,[6] Nobuo N. Noda,[7] Fuyuhiko Inagaki,[6] Toshiya Senda,[8,*] and Masanori Hatakeyama[1,*]

[1]Division of Microbiology, Graduate School of Medicine, University of Tokyo, Tokyo 113-0033, Japan
[2]Division of Chemistry, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan
[3]Structure Guided Drug Development Project, JBIC Research Institute, Japan Biological Informatics Consortium (JBIC), Tokyo 135-0064, Japan
[4]Division of Molecular Oncology, Institute for Genetic Medicine, Hokkaido University, Sapporo 060-0815, Japan
[5]National Institute of General Medical Sciences and National Cancer Institute Collaborative Access Team (GM/CA-CAT) at the Advanced Photon Source (APS), Bioscience Division, Argonne, National Laboratory, Argonne, Illinois 60439, USA
[6]Department of Structural Biology, Faculty of Advanced Life Science, Hokkaido University, Sapporo 001-0021, Japan
[7]Institute of Microbial Chemistry, Tokyo 141-0021, Japan
[8]Biomedicinal Information Research Center (BIRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan
*Correspondence: mhata@m.u-tokyo.ac.jp (M.H.), toshiya-senda@aist.go.jp (T.S.)
http://dx.doi.org/10.1016/j.chom.2012.05.010

SUMMARY

The *Helicobacter pylori* type IV secretion effector CagA is a major bacterial virulence determinant and critical for gastric carcinogenesis. Upon delivery into gastric epithelial cells, CagA localizes to the inner face of the plasma membrane, where it acts as a pathogenic scaffold/hub that promiscuously recruits host proteins to potentiate oncogenic signaling. We find that CagA comprises a structured N-terminal region and an intrinsically disordered C-terminal region that directs versatile protein interactions. X-ray crystallographic analysis of the N-terminal CagA fragment (residues 1-876) revealed that the region has a structure comprised of three discrete domains. Domain I constitutes a mobile CagA N terminus, while Domain II tethers CagA to the plasma membrane by interacting with membrane phosphatidylserine. Domain III interacts intramolecularly with the intrinsically disordered C-terminal region, and this interaction potentiates the pathogenic scaffold/hub function of CagA. The present work provides a tertiary-structural basis for the pathophysiological/oncogenic action of *H. pylori* CagA.

INTRODUCTION

Chronic infection with *Helicobacter pylori cagA*-positive strains is associated with atrophic gastritis and peptic ulcerations and is the strongest risk factor for the development of gastric carcinoma, the second leading cause of cancer-related death worldwide (Blaser and Atherton, 2004; Parsonnet et al., 1997). The *cagA*-encoded CagA is a 120∼145 kDa protein, which is delivered into gastric epithelial cells via a type IV secretion system (TFSS) (Covacci and Rappuoli, 2000). The CagA delivery requires binding of bacterial surface-exposed CagA, particularly at the TFSS pilus tip (Jiménez-Soto et al., 2009; Kwok et al., 2007), with phosphatidylserine (PS), which is normally concentrated to the inner leaflet of the plasma membrane in host cells but is transiently externalized to the membrane surface where *H. pylori* directly attached (Murata-Kamiya et al., 2010). Interaction of integrin β1 with CagA as well as TFSS components such as CagI, CagL and CagY is also involved in the translocalization of CagA (Jiménez-Soto et al., 2009; Kwok et al., 2007). The CagA protein is noted for sequence diversity in its C-terminal region that contains a variable number of Glu-Pro-Ile-Tyr-Ala (EPIYA) tyrosine-phosphorylation motifs, and, based on characteristics of the amino acids flanking each of the EPIYA motifs, four distinct EPIYA segments (A-D) have been identified (Higashi et al., 2002a; Xia et al., 2009). The C-terminal region of CagA from East Asian *H. pylori* strains (East Asian CagA) contains EPIYA-A and EPIYA-B segments, which are followed by one EPIYA-D segment, whereas that of *H. pylori* isolated from the rest of the world (Western CagA) contains EPIYA-A and EPIYA-B segments, followed by one to three repeats of EPIYA-C segment. Accordingly, variations in molecular size of CagA are mostly due to differential combinations of EPIYA segments.

Once inside the host cell, CagA is tethered to the inner face of the plasma membrane, where it exerts its pathophysiological action (Higashi et al., 2002b; Bagnoli et al., 2005). In nonpolarized epithelial cells, the EPIYA segment acts as a membrane-localization signal of CagA (Higashi et al., 2005). In contrast, in polarized epithelial cells, interaction of CagA with PS, which is abundantly present in the inner leaflet of the plasma membrane, plays an important role in the membrane association

of delivered CagA (Bagnoli et al., 2005; Murata-Kamiya et al., 2010). Membrane-tethered CagA then undergoes tyrosine phosphorylation at the EPIYA motifs by Src family kinases (SFKs) or c-Abl kinase (Backert and Meyer, 2006) and thereby acquires the ability to promiscuously interact with SH2-domain containing proteins such as SHP2 tyrosine phosphatase, C-terminal Src kinase (Csk), and Crk adaptor (Hatakeyama, 2004; Higashi et al., 2002b; Suzuki et al., 2005). Among these, SHP2 is known as a bona fide oncoprotein, gain-of-function mutations of which are associated with a variety of human malignancies (Mohi and Neel, 2007). Through complex formation, CagA also deregulates SHP2 phosphatase activity, causing aberrant activation of pro-oncogenic Erk MAP kinase signaling (Hatakeyama, 2004). Additionally, CagA binds to the polarity-regulating serine/threonine kinase Partitioning defective-1 (PAR1)/Microtubule affinity-regulating kinase (MARK) (Lu et al., 2009; Saadat et al., 2007; Zeaiter et al., 2008). The CagA-PAR1 interaction occurs between the 16 amino-acid residues termed the CM (CagA multimerization) or MKI (MARK-kinase inhibitor) sequence in C-terminal CagA and the kinase catalytic domain of PAR1 (Nesić et al., 2010; Ren et al., 2006; Saadat et al., 2007). Through this interaction, CagA inhibits PAR1 kinase activity and thereby causes junctional and polarity defects, which also contribute to epithelial cell transformation (Saadat et al., 2007). Conversely, PAR1, which forms a homodimer (Nagase et al., 2011), tethers two CagA proteins at the CM sequences and this CagA tethering potentiates CagA-SHP2 complex formation and subsequent deregulation of SHP2 (Lu et al., 2008, 2009; Saadat et al., 2007). Consequently, CagA-mediated signal perturbation triggers a morphogenetic change in gastric epithelial cells characterized by an extremely elongated cell shape known as the "hummingbird phenotype" (Segal et al., 1999). In vivo oncogenic potential of CagA has been demonstrated by the observation that transgenic CagA expression spontaneously induces gastrointestinal and hematological malignancies in mice (Ohnishi et al., 2008).

Elucidation of the three-dimensional structure of CagA is a major challenge that has so far been hampered by the large size, lack of homology, and high protease susceptibility of CagA (Angelini et al., 2009; Hohlfeld et al., 2006). In this work, we determined the tertiary structure of CagA, and we discuss it here in terms of CagA functions.

RESULTS

Biological and Physicochemical Properties of Recombinant *H. pylori* CagA Protein

The recombinant full-length CagA protein (a 135-kD Western CagA of 1186 amino-acid residues derived from *H. pylori* strain 26695) expressed in *Escherichia coli* as a GST fusion protein was purified using glutathione beads and was then cleaved with PreScission protease. Since CagA does not share sequence homology with any of the known proteins, we first wished to obtain structural information on CagA by limited proteolysis. Upon partial digestion with *Staphylococcus aureus* V8 protease, full-length CagA was cleaved into 100-kD and 35-kD fragments (Figure S1A). The V8 protease digestion pattern was reproducible with another CagA of *H. pylori* NCTC11637 strain origin (Figure S1B). Amino-acid sequencing revealed that the N terminus of the 100-kD fragment was identical to that of full-length CagA, whereas the 35-kD fragment started from residue 877 (L877) of CagA (Figure S1A). An immunoblotting experiment showed that the 35-kD fragment contained a hexahistidine (6xHis)-tag fused to the C terminus of recombinant CagA (Figure S1C). From these results, we denoted the N-terminal 100-kD fragment (residues 1-876) and the C-terminal 35-kD fragment (residues 877-1186) as CagA(1-876) and CagA(877-1186), respectively (Figure 1A). Cleavage of CagA into 100-kD and 35-kD fragments had been previously reported to occur (Jiménez-Soto et al., 2009; Moese et al., 2001), although the actual protease and pathobiological significance of this CagA processing is still not known. We also performed a limited proteolysis of recombinant CagA(1-876) with trypsin and obtained two major fragments, one corresponding to residues 1-260, CagA(1-260), and the other corresponding to residues 261-829, CagA(261-829) (Figures 1A, S1D, and S1E).

The C-terminal CagA(877-1186) contains the EPIYA segment and CM sequence, which respectively act as binding sites for SHP2 and PAR1 (Higashi et al., 2002b; Saadat et al., 2007). The scaffold/hub function of C-terminal CagA is responsible for the morphogenetic activity of CagA known as the hummingbird phenotype (Bagnoli et al., 2005; Higashi et al., 2005). To determine whether the recombinant CagA(877-1186) purified from *E. coli* was biologically active, we utilized CagA-PAR1 interaction as an experimental read-out that quantitatively evaluates the scaffold/hub function of C-terminal CagA. To do so, CagA(877-1186), either produced in *E. coli* or expressed in COS-7 cells, was mixed with lysates of COS-7 cells expressing FLAG-tagged PAR1b, a major PAR1 isoform in epithelial cells, and complex formation of CagA with PAR1b was examined by co-immunoprecipitation (Lu et al., 2008). The in vitro mixing experiment demonstrated that *E. coli*-derived CagA(877-1186) bound to PAR1b to a level comparable to that observed with COS-7-derived CagA(877-1186) (Figure 1B). We then tested the morphogenetic activity of recombinant CagA proteins purified from *E. coli* in AGS human gastric epithelial cells by microinjection and confirmed the hummingbird-inducing activity of CagA(877-1186), although the magnitude was less than that induced by full-length CagA (Figure 1C and Movies S1–S3). Thus, recombinant CagA(877-1186) had an attenuated but substantial level of CagA activity.

To investigate the physicochemical property of the biologically active C-terminal CagA, we performed ^{1}H-NMR analysis of recombinant CagA(877-1186). The ^{1}H-NMR spectrum showed weak dispersion and crowded peaks of amide proton chemical shifts around the 8-ppm region (Figure 1D). Spectrum data with similar characteristics were also obtained with recombinant CagA(873-1026), which corresponded to the EPIYA segments (EPIYA-A, -B and -C), purified from *E. coli* (Figure S1F). The results of the NMR study indicated that the CagA C-terminal fragment (residues 877-1186) is virtually free from a high-order structure. Since EPIYA segments of C-terminal CagA had been predicted to be disordered (Angelini et al., 2009; Nesić et al., 2010), our result extended results of previous studies by demonstrating that the CagA C-terminal region up to the very C terminus is intrinsically disordered and therefore lacks a solid structure.

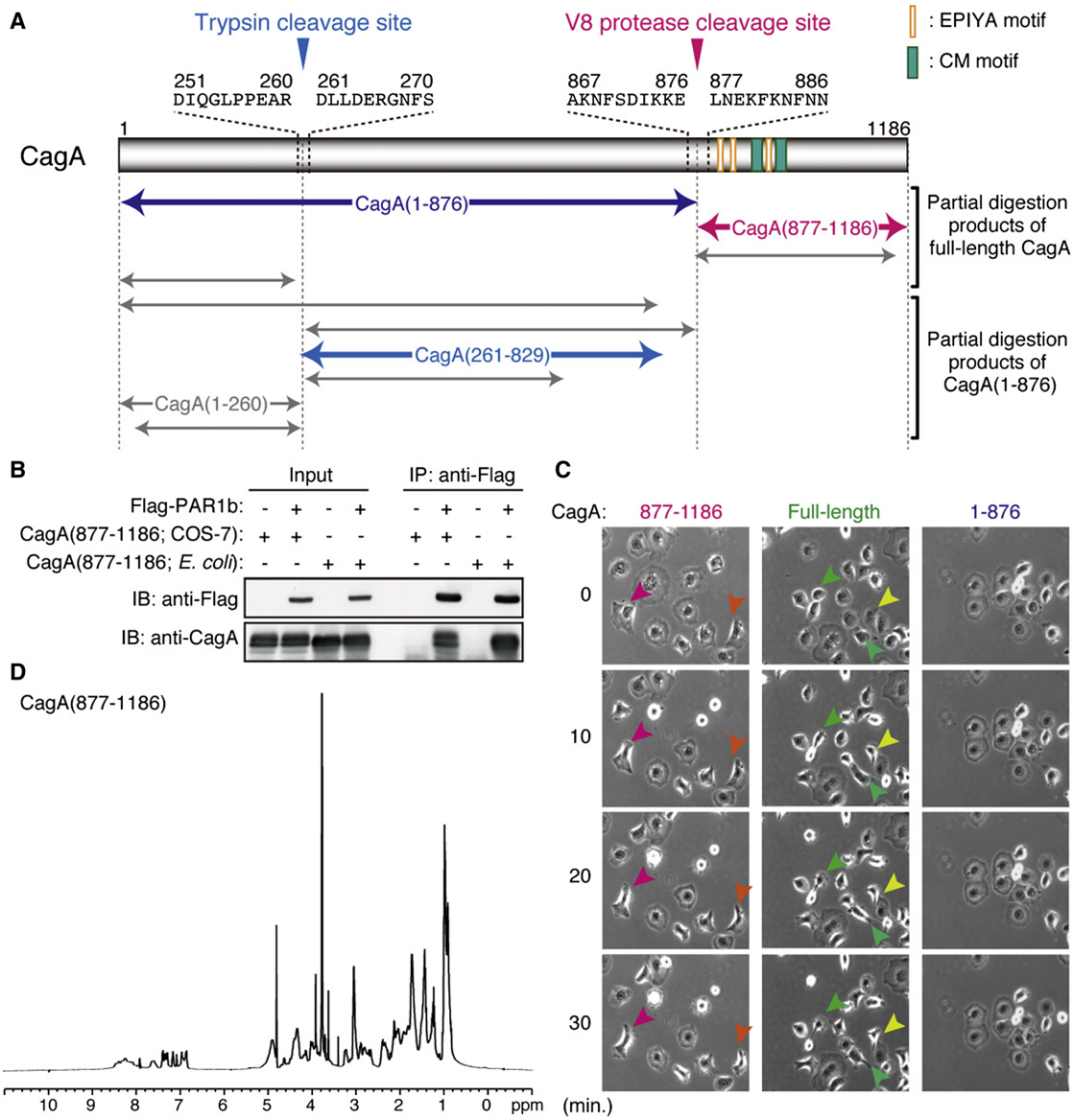


Figure 1. Structure and Activity of CagA Proteolytic Fragments

(A) Schematic representation of the partial digestion patterns of full-length CagA and CagA(1-876) by *S. aureus* V8 protease and trypsin.

(B) CagA(877-1186) produced in COS-7 cells or in *E. coli* was mixed with FLAG-tagged PAR1b expressed in COS-7 cells and the complex formation was determined by co-immunoprecipitation.

(C) Recombinant full-length CagA, CagA(1-876) or CagA(877-1186) purified from *E. coli* was microinjected into AGS cells and cell shapes were observed. All cells in the pictures were microinjected with CagA. Arrowheads indicate elongated cells.

(D) ^{1}H NMR spectrum of recombinant CagA(877-1186). See also Figure S1.

Intramolecular N-Terminal and C-Terminal Interaction of CagA

CagA binds to PAR1 via the C-terminal CM sequence (Lu et al., 2008; Nesić et al., 2010; Saadat et al., 2007). Using surface plasmon resonance (SPR) spectroscopy, we directly measured binding affinity of recombinant CagA to PAR1 (Figure 2A). Full-length CagA bound to PAR1b (residues 39-364) with a K_D value of 3.9 ± 0.64 nM. In contrast, C-terminal CagA(877-1186) bound to PAR1b (K_D = 33.0 ± 9.3 nM) with a strength that was one order of magnitude less than that of full-length CagA. Since N-terminal CagA(1-876) did not directly bind to PAR1, the result suggested the presence of a functional interplay between the N-terminal region and C-terminal region that strengthens binding of C-terminal CagA to PAR1. In this regard, Bagnoli et al. had reported that N-terminal CagA associates with C-terminal CagA and that this intramolecular interaction potentiates the morphogenetic activity of C-terminal CagA (Bagnoli et al., 2005). We therefore wished to elucidate the mechanistic details of the N-terminal/C-terminal interaction. To do so, we first narrowed down CagA sequences that mediate the intramolecular

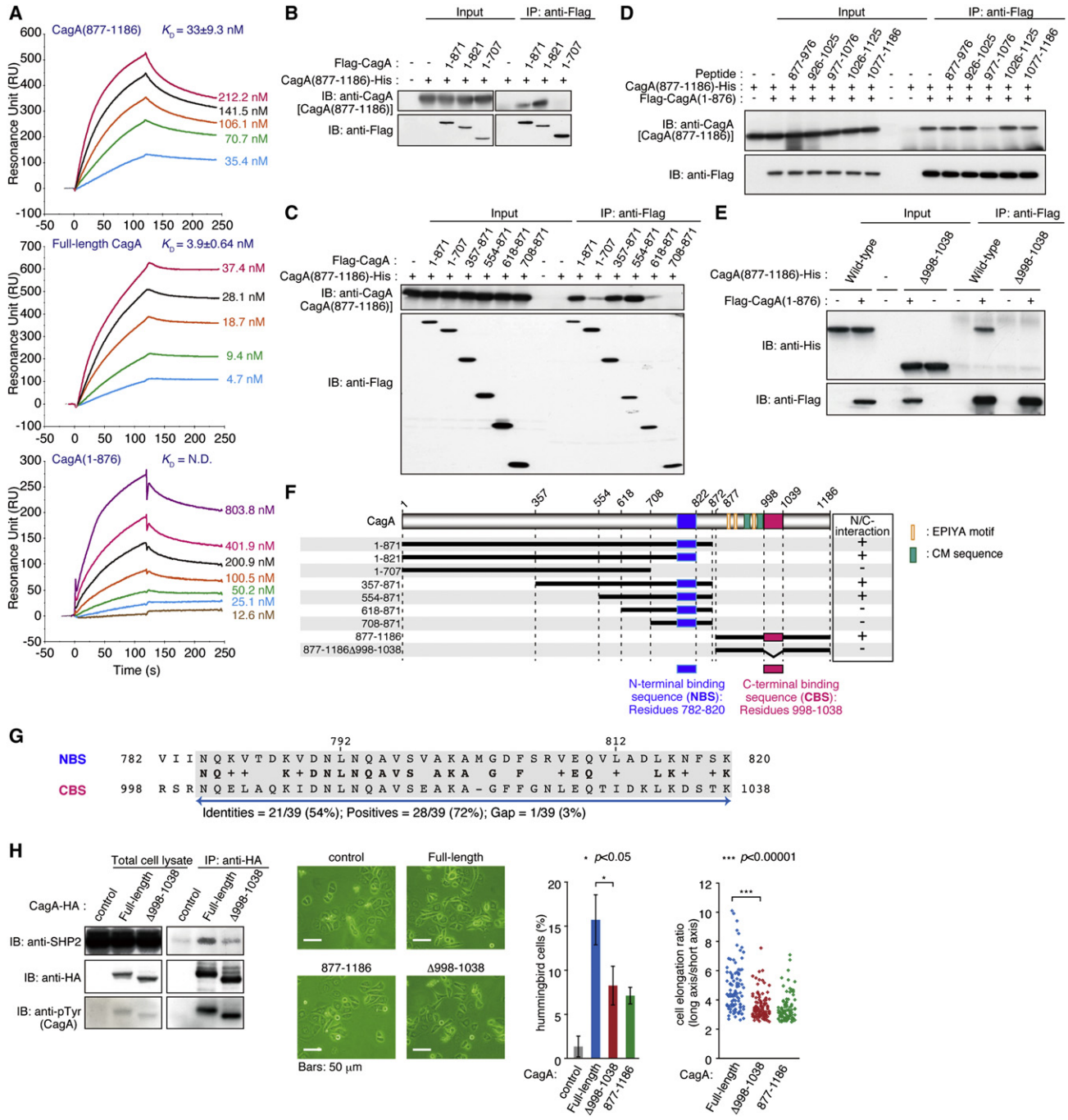


Figure 2. Intramolecular Interaction of CagA

(A) Binding affinities of recombinant CagA proteins to immobilized recombinant PAR1b (residues 39-364) were determined by SPR spectroscopy.

(B and C) Each of the FLAG-tagged CagA(1-876) mutants expressed in COS-7 cells was in vitro mixed with recombinant CagA(877-1186) and the complex formation was determined by co-immunoprecipitation.

(D) Interaction of recombinant CagA(877-1186) with FLAG-tagged CagA(1-876) expressed in COS-7 cells was examined by co-immunoprecipitation in the presence of the indicated peptide derived from C-terminal CagA.

(E) FLAG-tagged CagA(1-876) expressed in COS-7 cells was in vitro mixed with CagA(877-1186) or CagA(877-1186|Δ998-1038), in which residues 998-1038 were deleted from CagA(877-1186). The N-terminal/C-terminal interaction of CagA was then investigated by a co-immunoprecipitation experiment.

(F) Schematic view of CagA regions responsible for the intramolecular N-terminal/C-terminal intramolecular interaction.

(G) Amino-acid sequence homology between NBS and CBS.

(H) SHP2-binding activities of full-length CagA and CagAΔ998-1038 (left). AGS cells were transfected with the indicated CagA expression vectors (second left). The number of hummingbird cells was counted (second right). Relative cell length was also quantified (right). Error bars represent mean ± S.D., *p < 0.05 (Student's t test, n = 3), ***p < 0.00001 (Mann-Whitney *U* test, n = 100). See also Figure S2.

Table 1. Crystallographic Summary

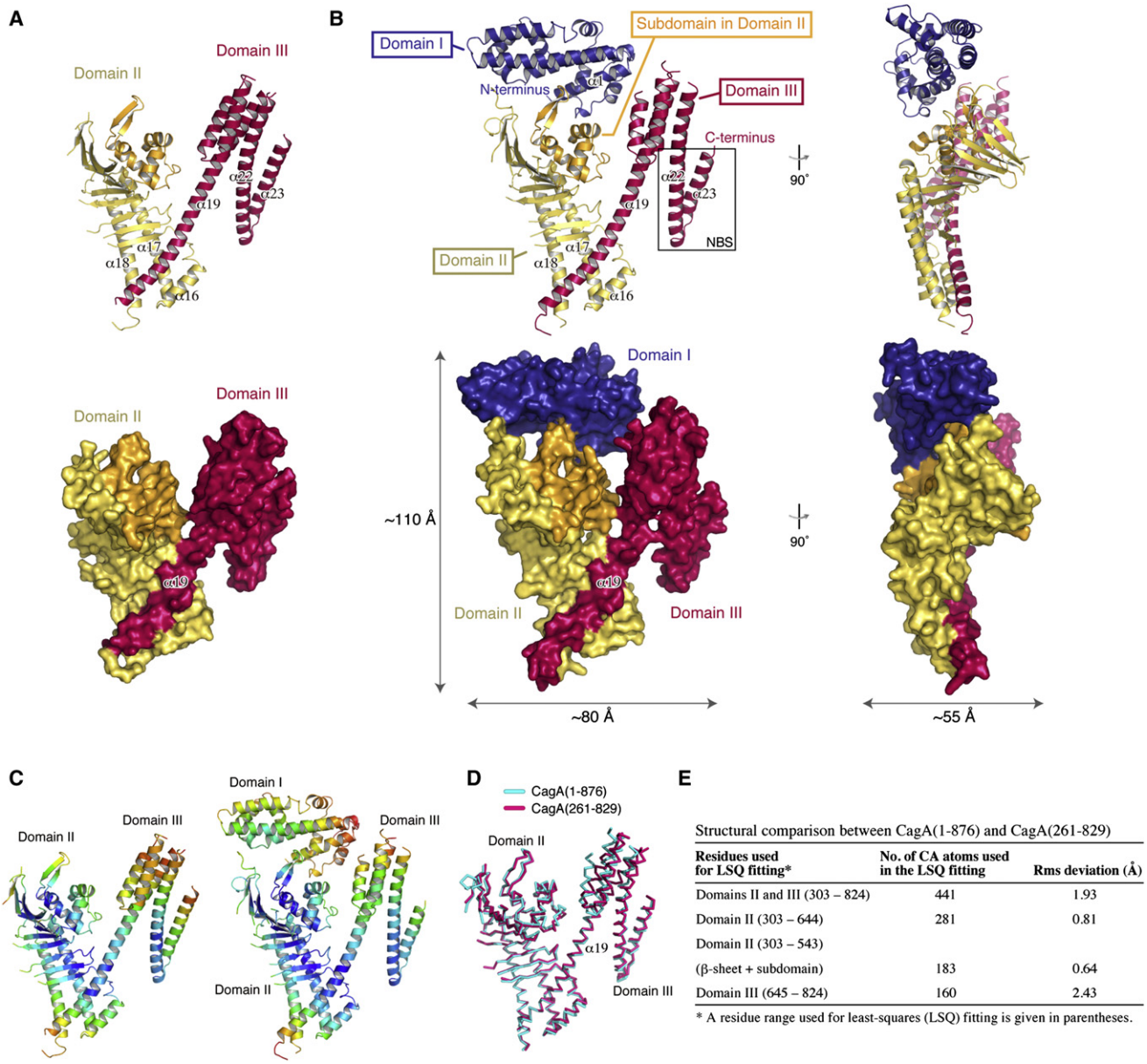
	MAD data (CagA(261–829))		CagA(261–829)	CagA(1–876)
Data Collection				
Space group	$P3_121$			$P4_12_12$
Cell dimensions				
a, *b*, *c* (Å)	96.41, 96.41, 171.59		95.67, 95.67, 167.92	97.31, 97.31 244.76
α, β, γ (°)	90, 90, 120		90, 90, 120	90, 90, 90
	Peak	*Inflection*		
Wavelength (Å)	0.9806	0.9809	0.9806	0.9790
Resolution (Å)	83.5 – 3.30 (3.47 – 3.30)	83.6 – 3.35 (3.53 – 3.35)	74.3 – 3.19 (3.37 – 3.19)	90.42 – 3.30 (3.58 – 3.30)
*R*merge	0.038 (0.420)	0.037 (0.427)	0.059 (0.468)	0.054 (0.5050)
$I/\sigma I$	30.70 (4.66)	32.51 (4.58)	33.39 (6.00)	25.23 (4.42)
Completeness (%)	99.2 (100)	99.4 (100)	99.3 (100)	99.5 (100)
Redundancy	5.7 (5.7)	5.7 (5.8)	10.6 (11.1)	7.6 (7.7)
Crystallographic Refinement				
Resolution (Å)			74.3 – 3.19 (3.41 – 3.19)	90.42 – 3.30 (3.50 – 3.30)
No. of reflections			15,242	18,403
*R*work/*R*free			0.2215/0.2786 (0.2478/0.3041)	0.1905/0.2476 (0.2241/0.2820)
No of atoms				
Protein			3635	5229
Ligand/Ion			-	-
Water			-	-
B-factors (Å^2)				
Protein			103.74	145.89
Ligand/Ion			-	
Water			-	
Rms deviations				
Bond lengths (Å)			0.010	0.010
Bond angles (°)			1.21	1.22

Each data set was collected from one crystal. Highest resolution shell is shown in parentheses.

interaction by in vitro mixing of lysates prepared from COS-7 cells expressing a series of CagA(1-876) mutants with recombinant CagA(877-1186). The results of a co-immunoprecipitation experiment revealed that deletion of residues 554-617 or 708-821 abolished the ability of N-terminal CagA to bind C-terminal CagA (Figures 2B and 2C). In addition, results of a binding-competition experiment using a series of peptides derived from C-terminal CagA and a co-immunoprecipitation experiment using internal deletion mutants of CagA(877-1186) indicated that residues 998-1038 are responsible for interaction with N-terminal CagA (Figures 2D–2F). To our surprise, the identified C-terminal binding residues 998-1038 exhibited significant sequence homology (54% identity) with the N-terminal residues 782-820, which were within the N-terminal CagA sequence required for C-terminal CagA binding (Figure 2G). The homology was more evident when nucleotide sequences were compared (70% identity) (Figure S2), suggesting that these two CagA sequences have been derived from the same ancestral DNA fragment through duplication. We termed residues 782-820 as N-terminal binding sequence (NBS) and residues 998-1038 as C-terminal binding sequence (CBS) (Figure 2F). Whereas CBS contained neither an EPIYA motif nor CM sequence, a CagA mutant lacking CBS (CagAΔ998-1038) had substantially reduced SHP2-binding activity and decreased hummingbird-inducing activity in AGS cells (Figure 2H), providing additional evidence for the importance of intramolecular CagA interaction in enhancing the pathophysiological activity of C-terminal CagA. Given that CagA-PAR1 interaction is important for stable CagA-SHP2 complex formation (Saadat et al., 2007), the attenuated CagA activity was most probably due to weakened CagA-PAR1 interaction in the absence of N-terminal/C-terminal interaction of CagA.

Crystal Structure of the CagA N-Terminal Region

Next, we sought to crystallize the CagA protein. Whereas recombinant CagA(1-876) and CagA(261-829) were successfully crystallized, full-length CagA crystal was never obtained, suggesting that the disordered C-terminal CagA hampered crystallization of full-length CagA. X-ray analysis initially solved the crystal structure of CagA(261-829) at 3.19 Å resolution by the multiwavelength anomalous dispersion (MAD) method (Table 1 and Figure 3A). To validate sequence assignment of the crystal structure, we prepared 31 mutant derivatives, each of which contains different selenomethionine (SeMet) substitutions (Table S1 and Figure S3A). Using the CagA(261-829) crystal structure as a search model, the crystal structure of CagA(1-876) was



E

Structural comparison between CagA(1-876) and CagA(261-829)

Residues used for LSQ fitting*	No. of CA atoms used in the LSQ fitting	Rms deviation (Å)
Domains II and III (303 – 824)	441	1.93
Domain II (303 – 644)	281	0.81
Domain II (303 – 543) (β-sheet + subdomain)	183	0.64
Domain III (645 – 824)	160	2.43

* A residue range used for least-squares (LSQ) fitting is given in parentheses.

Figure 3. Crystal Structure of the CagA N-Terminal Region

(A and B) Cartoon diagram (upper) and surface representation (lower) of the crystal structures of CagA(261-829) (A) and CagA(1-876) (B).
(C) B-factor representation of CagA(261-829) (left) and CagA(1-876) (right) from blue (low) to red (high).
(D) Domains II and III in CagA(261-829) (pink) and CagA(1-876) (cyan) were superimposed by least-squares fitting.
(E) Summary of structural comparison between CagA(1-876) and CagA(261-829). See also Figure S3.

next determined by the molecular replacement (MR) method (Table 1 and Figure 3B). Since the electron density of CagA(1-876) was poor, 10 mutants of CagA(1-876) were utilized for amino acid assignment (Table S1 and Figure S3A). The electron density map obtained from the CagA(1-876) crystal resolved the N terminus end region of CagA (residues 24-221), which corresponded to the smaller fragment of the trypsin-digested CagA N-terminal region (Figure 1A). On the other hand, no electron densities were observed for residues 825-876, suggesting that these residues were structurally disordered. We thus concluded that the entire CagA protein comprises a structured N-terminal region (residues 1-829) (Table S2) and an intrinsically disordered C-terminal region (830-1186).

The crystal structure of the CagA N-terminal region (residues 1-829) assumed a square plate-like shape with approximate dimensions of 110 × 80 × 55 $Å^3$ (Figures 3B and 4). Whereas the overall structure was defined, N-terminal CagA was characterized by the existence of short disordered stretches such as residues 222-302, 479-488, and 510-536, of which electron densities were not observed in any crystal form (Figure 4 and

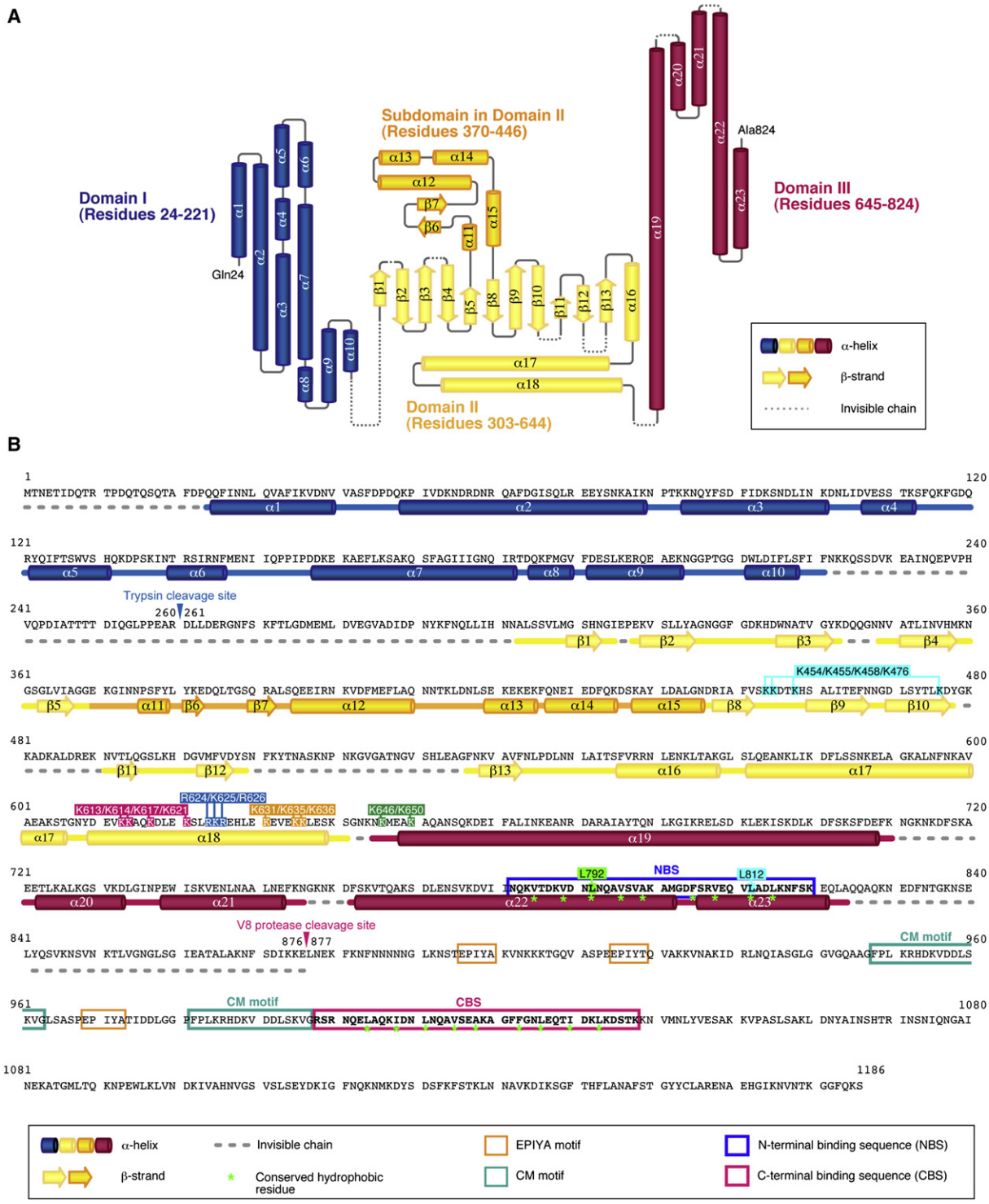


Figure 4. Domain definition and sequence alignment with secondary structure elements of CagA

(A) Topological configurations of the secondary structures of CagA(1-876).

(B) The secondary structure was assigned to CagA sequence based on the crystal structure of CagA(1-876). See also Figure S4.

Table S2). N-terminal CagA consists of three domains, termed Domains I-III. Domain I, the most N-terminal domain, was made up of 10 α helices (α1–α10) and was structurally isolated from the other two domains; Domain I has a small interacting surface area (374 Å^2) with Domain II but has no interaction with Domain III. Because of this weak interaction with other parts of

the molecule, Domain I is likely to have a mobile character, which was supported by high B-factors of residues in Domain I (Figure 3C). Domain II and Domain III, which corresponded to residues 303-644 and 645-824 (Figure 4), respectively, comprised a protease-resistant CagA structural core that showed an N-shaped dimodular architecture. Domains II and III were connected by a long helix α19, which is kinked at around residue 670 by approximately 30°. Domain II contained a large antiparallel β sheet, which had an insertion of a subdomain (residues 370-446) between β5 and β8 strands. This subdomain, located at the center of the N-terminal region of CagA, tightly interacted with the inner surface of the β sheet. Two α helices, α17 and α18, also interacted with the inner surface of the β sheet. B-factor values of the residues in the large β sheet and the subdomain were lower than those of residues in other regions (Figure 3C), suggesting that these structural elements comprise a rigid core of CagA. Indeed, this region showed a small structural deviation between CagA(1-876) and CagA(261-829); the root-mean-square structural displacement was 0.64 Å for 183 CA atoms (Figures 3D and 3E). On the other hand, the long α19 helix that connects Domain II and Domain III seemed to be flexible (Figure S3C). The bent angle of the α19 helix was significantly different between CagA(1-876) and CagA(261-829), resulting in a change of relative disposition of Domain III between CagA(1-876) and CagA(261-829) (Figure 3D). In Domain III, helices α20, α21, and α22 formed a 4-helix bundle with the C-terminal region of the α19 helix (Figures 3A, 3B, and 4). Residues 554-617 and residues 708-821, deletion of which abolished the intramolecular CagA interaction, constituted helices α16-α18 in Domain II and helices α20-α23 in Domain III, respectively. Also, NBS (residues 782-820), which associates with CBS (residues 998-1038) in C-terminal CagA, formed a part of helices α22/23 (Figure 4B). A structural homology search using the DALI server (Holm and Sander, 1996) revealed that the tertiary structure of the N-terminal CagA region was unprecedented, with no homologous structure reported (Figure S4).

Basic Amino-Acid Cluster in Domain II

CagA binds to the acidic membrane phospholipid, PS (Murata-Kamiya et al., 2010). Whereas PS is physiologically concentrated to the inner leaflet of the plasma membrane, it is transiently externalized to the outer leaflet at the membrane site where *H. pylori* directly attached. Interaction of CagA with PS is important not only for the delivery of CagA into host cells but also for the localization of delivered CagA to the inner face of the plasma membrane, especially in polarized epithelial cells (Murata-Kamiya et al., 2010). Two arginine residues, R624 and R626 (R619 and R621 in NCTC11637-derived CagA), both of which are crucial for the CagA-PS interaction, were located in helix α18 of Domain II (Figure 5A). The crystal structure of N-terminal CagA revealed that, together with R624 and R626, lysine residues at 454, 455, 458, 476, 613, 614, 617, 621, 625, 631, 635, 636, 646, and 650 constitute a basic amino-acid cluster that provides a positive electrostatic surface potential (Figure 5A). Given this, we hypothesized that the basic patch mediates the CagA-PS interaction. A series of basic patch mutants of CagA were then prepared and subjected to an in vitro lipid-binding assay or transient expression in polarized Madin-Darby canine kidney (MDCK) cells. The results of the experiments revealed involvement of K613, K614, K617, K621, K631, K635 and K636, in addition to R624 and R626, in the CagA-PS interaction (Figures 5B and 5C), confirming that the basic patch acts like Velcro to mediate the CagA-PS interaction. From this observation, we also assumed that the basic patch faces the inner side of the plasma membrane, where PS is present, and deduced molecular orientation of delivered CagA relative to the plasma membrane as described in Figure 5D.

Functional Studies of the Structured CagA N-Terminal Region

Packing analysis of the $P4_12_12$ and $P3_121$ crystals showed that similar intermolecular interactions that dimerize N-terminal CagA occurred in both crystal forms (Figures 6A and S5A). The two molecules in the dimer were related to each other by a crystallographic two-fold axis in both crystal forms; helices α22 and α23 of two adjacent N-terminal CagA fragments underwent hydrophobic interactions by forming an intermolecular 4-helix bundle in these dimers. In the $P4_12_12$ crystal, the dimerization interface comprised nine hydrophobic residues, centering around L792 in helix α22 and L812 in helix α23. (Figure 6A). The interaction buried 627 $Å^2$ of the surface per partner molecule. Substitution of L792 or L812 with SeMet caused significant loss of integrity of the mutant crystal (Table S1). The N-terminal dimerization and the formation of intermolecular 4-helix bundle were also preserved in the CagA(261-829) crystal with the $P3_121$ space group (Figure S5A), although relative arrangement of the two molecules was slightly different from that observed in the $P4_12_12$ crystal. These observations indicated the importance of intermolecular contacts for the quality of CagA crystals. They also raised the idea that N-terminal CagA spontaneously dimerizes independently of C-terminal CagA. However, the results of static light scattering analysis demonstrated that full-length CagA or the CagA(1-876) fragment purified from *E. coli* was present as a monomer in solutions up to the concentration of 6 μM (Figure 6B). In this regard, the first peak of full-length CagA in the gel filtration profile corresponded to a CagA dimer. However, the peak was markedly decreased in the presence of dithiothreitol (Figure S5B). Also, such dimer peak was absent when the N-terminal CagA(1-876) fragment was applied (Figure 6B). These results indicated that the dimer peak was artificially made by an intermolecular disulfide bond at C1164, the only cysteine residue present in C-terminal CagA (residues 877-1186). Furthermore, a GST pull-down experiment using a solution containing 72 μM GST-CagA(1-876) and 310 μM FLAG-CagA(1-876) provided no evidence for the spontaneous dimerization of N-terminal CagA (data not shown). Together with the small interaction surface of the dimer (627 $Å^2$) in the crystal, these results favored the idea that the N-terminal CagA dimerization was due to crystal packing. Although the biological relevance of the N-terminal CagA dimerization remains unknown, we noticed that the sequence that forms the dimer interface in the crystal overlaps with NBS (residues 782-820) (Figure 2F). Furthermore, nine hydrophobic residues in NBS that constitute the hydrophobic interaction surface in the crystal were completely conserved in the NBS-related CBS (Figure 4B). This unexpected finding led us to examine whether the N-terminal dimerization interface in the crystal was also utilized for the intramolecular N-terminal/C-terminal interaction. To this

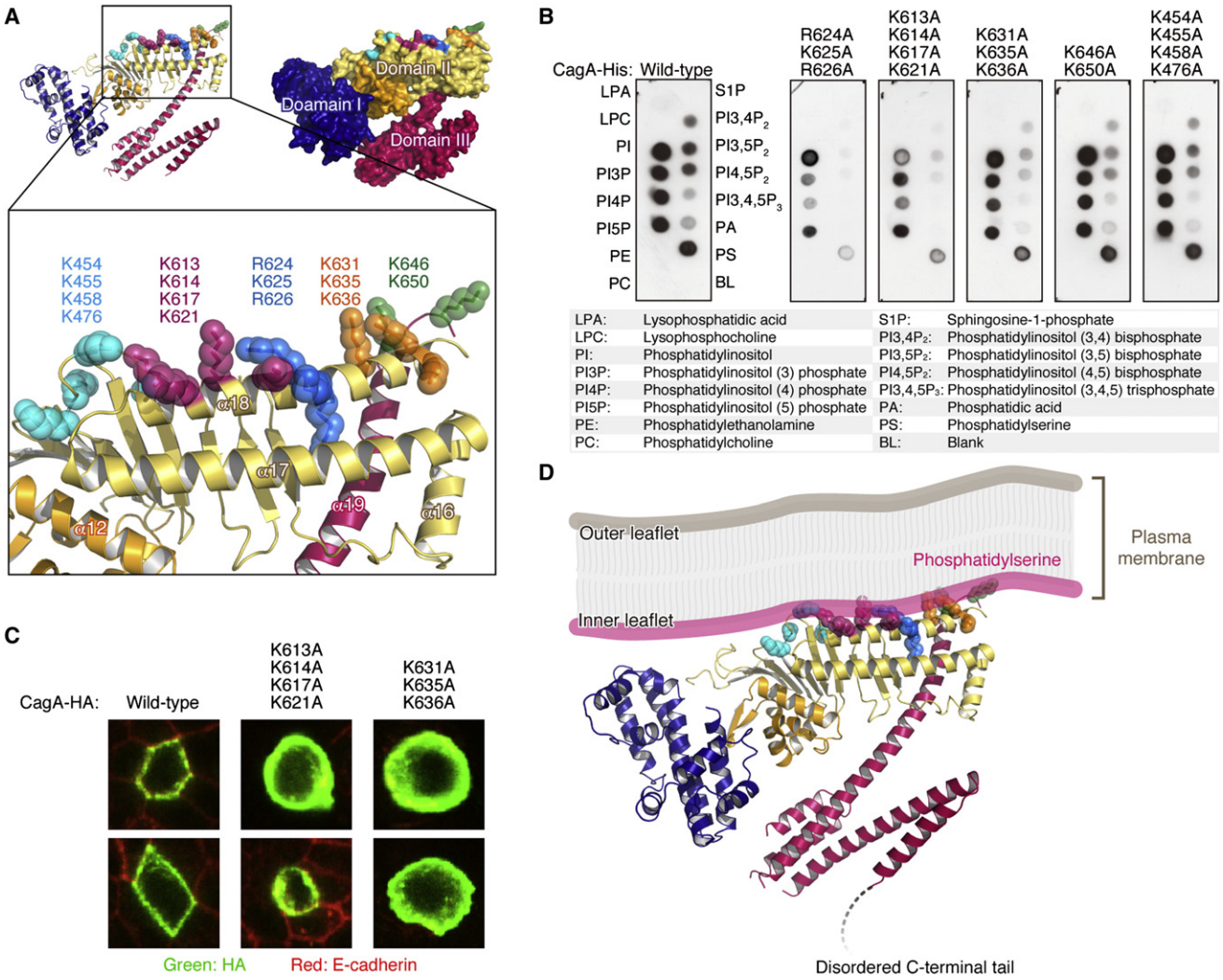


Figure 5. Involvement of Basic Patch in the Membrane Association of CagA

(A) Clustered amino-acid residues that comprise the basic patch on the Domain II of N-terminal CagA are highlighted in different colors.

(B) Phospholipid-binding assay of recombinant CagA proteins having variable combinations of mutations in the basic patch.

(C) Polarized MDCK cells were transfected with the indicated CagA vector, and subcellular localization of CagA was investigated by immunofluorescence microscopy.

(D) A model of CagA-host membrane interaction.

end, we prepared GST-fused CagA(1-876) and its derivatives, in which one or both of L792 and L812 were replaced by arginine. GST pull-down experiments using a mixture of the GST-fusion protein and C-terminal CagA(877-1186) revealed that the N-terminal/C-terminal interaction was completely abolished by L792R or L812R mutation in NBS (Figure 6C), indicating a critical role for the hydrophobic NBS/CBS interaction in the N-terminal/C-terminal association of CagA. In this regard, N-terminal deletion to residue 617 also abolished the N-terminal/C-terminal interaction (Figures 2C and 2F). Whereas the deletion does not involve NBS, deletion of most part of Domain I and Doman II seems to affect Domain III, which most likely disintegrates the CagA structure required for the NBS-CBS interaction.

Since CBS was located immediately distal to the PAR1-binding CM sequence in C-terminal CagA (Figure 4B), it was possible that the intramolecular NBS/CBS interaction influences the binding of PAR1 to the CM sequence. By surface plasmon resonance analysis, the equilibrium dissociation constants (K_D) obtained for binding of PAR1 to wild-type CagA and CagA(L812R) were almost the same. However, the kinetics of PAR1 binding to CagA(L812R) were substantially different from those obtained using wild-type CagA: PAR1 associated with and dissociated from CagA(L812R) more rapidly than to wild-type CagA (Figure 6D, lower). The results indicated that turnover of the CagA-PAR1 complex was accelerated in the absence of the intramolecular NBS/CBS interaction of CagA. Since PAR1-mediated tethering of two CagA proteins via the CM sequence markedly stimulates CagA-SHP2 complex formation (Saadat et al., 2007), elevated turnover of the CagA-PAR1 complex was thought to influence the CagA-SHP2 interaction. Indeed,

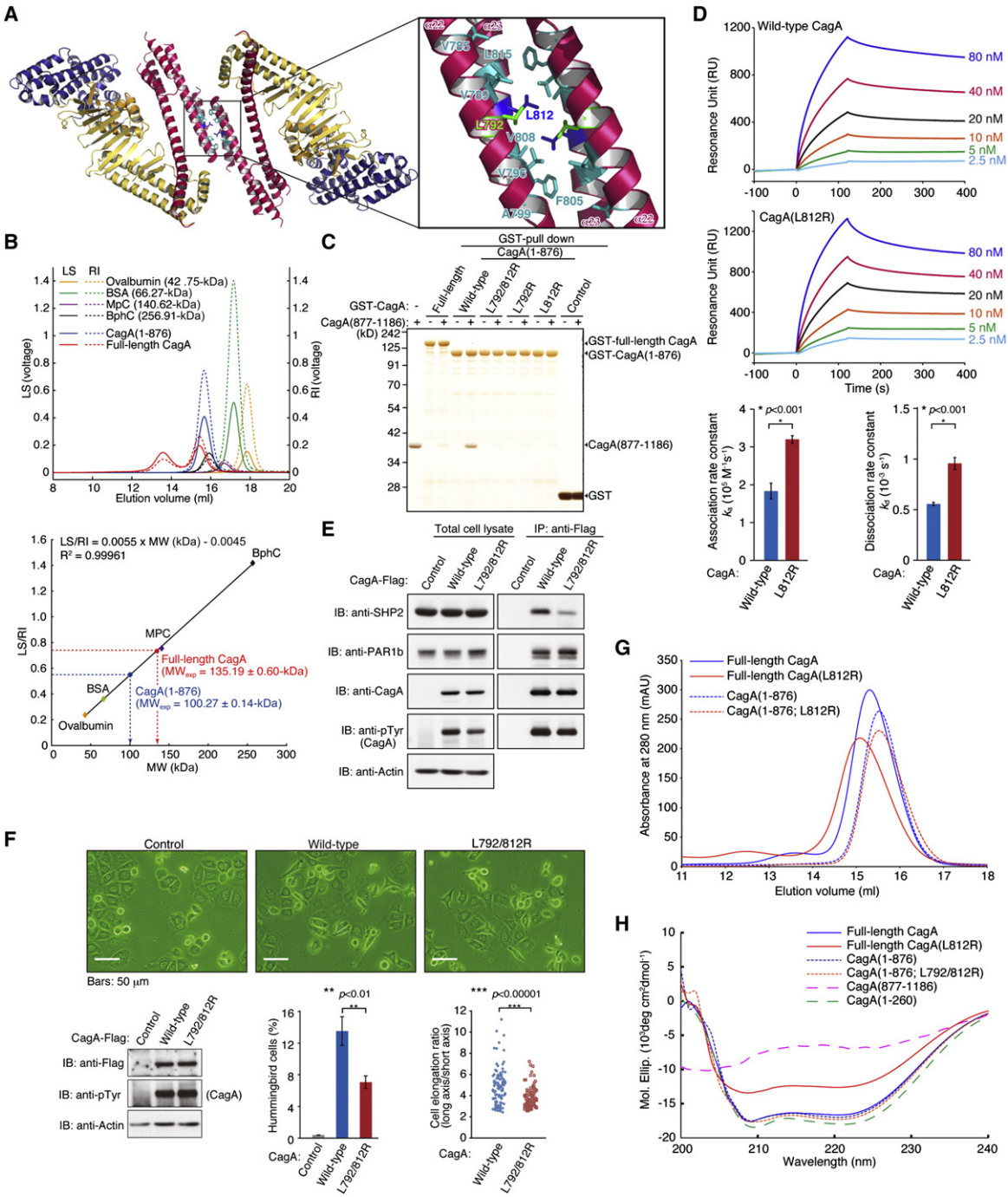


Figure 6. Role of the N-Terminal Dimerization Surface in the N-Terminal/C-Terminal CagA Interaction

(A) A CagA(1-876) dimer in the crystal. An intermolecular 4-helix bundle is formed between two adjacent CagA(1-876) monomers.

(B) Molecular weights of CagA in solution were determined by measuring intensity of static light scattering (LS) and refractive index (RI).

(C) CagA(877-1186) mixed with GST-fused full-length CagA or CagA(1-876) was subjected to a GST pull-down assay.

(D) Binding kinetics between PAR1b and CagA mutant that cannot undergo intramolecular NBS/CBS interaction. Error bars represent mean ± S.D., *p < 0.001 (Student's t test, n = 3).

(E) Lysates of AGS cells transfected with the indicated FLAG-CagA vector were immunoprecipitated (IP) with an anti-FLAG antibody, followed by immunoblotting (IB) with the respective antibodies.

(F) Induction of hummingbird cells in AGS cells by CagA (upper) was statistically analyzed (lower). Error bars represent mean ± S.D., **p < 0.01 (Student's t test, n = 3), ***p < 0.00001 (Mann-Whitney *U* test, n = 100).

(G) Gel filtration profiles of recombinant CagA proteins.

(H) Prediction of secondary structure contents for recombinant CagA proteins purified from *E. coli* by CD spectra. See also Figure S5.

CagA mutants that did not undergo NBS/CBS interaction exhibited reduced complex formation with SHP2 and showed decreased hummingbird-inducing ability in AGS cells (Figures 6E, 6F, S5C, and S5D). Next, to gain insights into the mechanism that affects the stability of the CagA-PAR1 complex, we investigated structural alterations in the disordered C-terminal region following the NBS/CBS interaction. Notably, the intramolecular interaction creates a lariat-like C-terminal loop (Figure S5E) while making the entire CagA structure more compact (Figure 6G). By circular dichroism (CD) spectrum analysis of *E. coli*-purified recombinant proteins, the N-terminal CagA(1-876) fragment gave particular CD peaks at 208 nm and 222 nm derived from α helices, and the helix content was estimated from CD spectra data to be approximately 58%, whereas the C-terminal CagA(877-1186) was virtually free of a secondary structure (Figure 6H). To our surprise, the CD spectra of full-length CagA were similar to those of CagA(1-876) despite the presence of the disordered C-terminal CagA region. The helix content of full-length CagA was, however, reduced to 41% upon introduction of the L812R mutation that disabled the NBS/CBS interaction (Figure 6H). Since a stronger L792/812R mutation did not alter the helix content of the CagA(1-876) fragment (Figure 6H), the possibility that the L812R mutation perturbed the overall structure of the folded N-terminal CagA was excluded. These results collectively indicated that, upon NBS/CBS interaction, a substantial portion of the intrinsically disordered C-terminal CagA adopted a secondary structure known as "induced folding" or "disorder-to-order transition," which plays an important role in the interaction of intrinsically disordered protein/region with target protein (Dunker et al., 2008; Dyson and Wright, 2005; Fink, 2005; Sigalov, 2010). Adaptive folding induced within or near the CM sequence in the C-terminal loop may therefore underlie reduced turnover of the CagA-PAR1 complex in response to the intramolecular NBS/CBS interaction, which potentiates the pathogenic scaffold/hub function of CagA.

DISCUSSION

In the present study, we elucidated the tertiary structure of the *H. pylori* CagA protein, the only known bacterial oncoprotein playing a critical role in human gastric carcinogenesis. CagA consists of a structured N-terminal region (~70% of the entire protein) and an intrinsically disordered C-terminal region (~30% of the entire protein). The crystal analysis revealed that the structured N-terminal CagA comprises three domains, Domains I-III. Domain I, which comprises the CagA N-terminal end, is linked to Domain II by a disordered region with about 80 amino acids. Since interaction surface is very small (385 Å^2), Domain I could be easily dissociated from Domain II and thus could be highly mobile in solution. Domain II and Domain III, bridged by a single long α-helix, form a unique "N-shaped" structural core of CagA made from 13 α helices and a huge β sheet. Interestingly, the structured N-terminal CagA still contains short disordered regions such as residues 222-302, 479-488, and 510-536. Since these disordered regions are common in all crystal forms, their presence may be a unique structural property of N-terminal CagA. Such disordered regions would not only act as flexible linkers between structured portions/domains but also serve as sites for molecular interaction, modification and so on. Indeed, N-terminal residues 204-236, which form a short helix and a disordered segment, have been reported to interact with RUNX3, a gastric tumor suppressor (Tsang et al., 2010). Residues 120-152 located on the same side of Domain I then promote degradation of CagA-bound RUNX3 by recruiting E3 ubiquitin ligase. This RUNX3 degradation was presumed to be mediated via WW domains in these CagA sequences. Our structural data, however, argue against the presence of such domains in N-terminal CagA.

In the crystal, N-terminal CagA dimerizes by hydrophobic interaction, in which L792 in helix α22 and L812 in helix α23 act as critical residues. Whereas the observed N-terminal CagA dimerization could be due to crystal packing, these two leucine residues within NBS were also indispensable for the intramolecular interaction of N-terminal CagA with C-terminal CagA, which appears to be mediated through the hydrophobic interaction between NBS and CBS (Figure S5E). Intriguingly, CBS exhibits substantial homology with NBS, most probably because these two sequences have been derived from an ancestral genomic segment. Despite the sequence similarity, however, NBS is folded into helices, while CBS is intrinsically disordered. We suspect that, upon interaction with NBS, CBS undergoes an induced folding that resembles the NBS helices. Extending this idea, the intermolecular NBS-NBS interaction in N-terminal CagA crystal, mediated by 4-helix bundle formation, might reflect the intramolecular interaction between NBS and NBS-related CBS in full-length CagA (Figure S5E). This idea was supported by 4-helix bundle formation between NBS and CBS through molecular modeling study (Figure S5F).

H. pylori injects CagA into the attached cells via TFSS. Although it remains unknown if CagA is delivered through the TFSS channel, the calculated size of the structured N-terminal CagA (110 × 80 × 55 Å^3) is smaller than the estimated diameter for the conduit of the *H. pylori* TFSS pilus (120 Å) (Rohde et al., 2003). Upon delivery, CagA is tethered to the inner face of the plasma membrane by interacting with PS, which is selectively enriched in the inner leaflet (Murata-Kamiya et al., 2010). A cluster of basic residues (basic patch) present on the surface of Domain II acts like Velcro to direct the CagA-PS interaction (Figure S5G). Involvement of multiple lysine and arginine residues in the CagA-PS interaction suggests the role of a polyelectrostatic effect in which the net charge of the basic patch influences the strength of CagA binding to PS. The CagA N-terminal residues 1-200 have recently been reported to act as yet another membrane-binding site, which targets CagA to the different membrane compartments and thereby regulates CagA functions (Pelz et al., 2011).

CagA exerts its oncogenic action by promiscuously interacting with multiple host proteins (Hatakeyama, 2008; Selbach et al., 2009). The intrinsically disordered nature of C-terminal CagA enables versatile CagA interaction because of its structural flexibility (Dunker et al., 2008; Dyson and Wright, 2005; Fink, 2005; Sigalov, 2010). This promiscuous CagA interaction is in many cases mediated via the EPIYA segment or CM sequence in the disordered C-terminal CagA (Hatakeyama, 2008). The EPIYA segments bind to SH2 domain-containing proteins such as SHP2, Csk, Grb2 and Crk (Hatakeyama, 2004; Mimuro et al., 2002; Selbach et al., 2009), whereas the CM sequence binds to PAR1 and c-Met receptor (Nesić et al., 2010; Saadat et al.,

2007; Suzuki et al., 2009). NBS located on the surface of helices α22 and α23 serves as a regulatory sequence that alters structural and dynamic characteristics of C-terminal CagA by interacting with CBS. Especially, the intramolecular NBS/CBS interaction influences turnover of the CagA-PAR1 complex formed via the CM sequence (Figure 6D). Of note, the CM sequence is localized immediately distal to CBS and thus the NBS/CBS interaction creates a lariat-like loop containing CM and EPIYA sequences in the disordered C-terminal CagA. Because the loop is fixed at both ends on the folded N-terminal CagA, the loop formation makes the entire CagA structure more compact, while C-terminal CagA becomes less mobile and less rotatable. The conformational constraints may trigger disorder-to-order transition observed in C-terminal CagA (Figure 6H), which underlies reduced turnover of the CagA-PAR1 complex that potentiates CagA-SHP2 interaction and subsequent induction of the hummingbird phenotype. The loop formation also fixes relative positioning of C-terminal CagA to N-terminal CagA, which might simplify tethering of two CagA proteins by a PAR1 dimer (Figure S5G) that could additionally promote CagA-SHP2 interaction (Saadat et al., 2007; Hatakeyama, 2008). The NBS/CBS interaction may therefore facilitate promiscuous interaction of CagA with target proteins by inducing adaptive structures in the disordered C-terminal CagA. This in turn indicates that N-terminal CagA serves as a positive regulatory element of C-terminal CagA by enhancing CagA membrane localization via the basic patch and by strengthening the pathogenic scaffold/hub function of CagA through the NBS/CBS interaction. This idea is also supported from the structural point of view that NBS is localized to the opposite side of the basic patch in the folded N-terminal CagA (Figure S5G).

Recent studies have also suggested additional roles of N-terminal CagA in the function and regulation of CagA. Beside its interaction with the above-described RUNX3 tumor suppressor, the CagA N-terminal region associates with apoptosis-stimulating protein of p53 (ASPP2) and thereby promotes p53 degradation (Buti et al., 2011). N-terminal CagA may therefore contribute to gastric carcinogenesis by inhibiting tumor suppressor proteins in host cells. Moreover, N-terminal CagA is required for the delivery of CagA into host cells by the *H. pylori* TFSS system (Hohlfeld et al., 2006) and is involved in the translocalization of CagA across the host cell plasma membrane by interacting with integrin β1 (Jiménez-Soto et al., 2009).

Elucidation of the tertiary structure of *H. pylori* CagA should enable comprehensive understanding for the promiscuous/versatile interaction of CagA with bacterial and host proteins in a spatiotemporal manner. The present study also opens the way to the development of a molecular targeting strategy that effectively inhibits the oncogenic action of CagA.

EXPERIMENTAL PROCEDURES

Protein Expression, Purification, and Crystallization

Protocols used for DNA construction, protein expression in *E. coli*, protein purification, and crystallization are outlined in the Supplemental Experimental Procedures.

Limited Proteolysis

Recombinant CagA proteins were incubated with *S. aureus* V8 protease or trypsin. The protein fragments produced by partial digestion were subjected to SDS-PAGE and subsequently silver-stained or electro-blotted onto a PVDF membrane. After coomassie brilliant blue staining of the PVDF membrane, the visualized protein bands were cut out and the N-terminal sequences were determined by a peptide sequencer.

Circular Dichroism (CD) Spectropolarimetry

CD spectra of recombinant CagA proteins were measured in 12.5 mM Tris-HCl, pH 7.0, 25 mM NaCl, 0.25 mM EDTA, and 0.25 mM DTT at 20°C using Jasco J-725 and J-820 spectropolarimeters.

^{1}H-NMR Spectroscopy

^{1}H-NMR spectra of recombinant CagA proteins were measured in 50 mM Tris-HCl, pH 7.0, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT at 25°C using a Varian UNITY INOVA 500 spectrometer. 2,2-Dimethylsilapentane-5-sulphonic acid was added to the protein sample and used as a reference standard.

Crystal Structure Determination

Diffraction data of CagA crystals were collected at PF-AR NE3A of PF in KEK (Tsukuba, JAPAN) and 23-ID-D of the Advanced Photon Source (Argonne, USA). The crystal structure of CagA(261–829) was determined by the MAD method (Table 1) using the program autoSHARP (Vonrhein et al., 2007). Density modification was performed using the program SOLOMON (Abrahams and Leslie, 1996). The molecular model was constructed using the program Coot (Emsley et al., 2010). Crystals of selenomethionine (SeMet)-introduced CagA point mutants were utilized to help amino-acid assignment of the molecular model (Table S1). The crystal structure of CagA(261-829) was refined using programs Phenix.refine (Adams et al., 2010) and BUSTER (Global Phasing Ltd., Cambridge, UK) (Table 1). The crystal structure of CagA(1-876) was determined by the MR method using autoMR in Phenix (Adams et al., 2010) and refined using programs Phenix.refine and BUSTER (Global Phasing Ltd., Cambridge, UK) (Table 1). In the structures of CagA(261-829) and CagA(1-876), no residues have $\phi-\psi$ angles in the disallowed and generously allowed regions of their Ramachandran plots (Figure S3B). Details of the crystallographic analysis are given in the Supplemental Experimental Procedures.

Phospholipid-Binding Assay

Phospholipid-binding assay on a membrane filter was performed using recombinant CagA proteins. Details of the assay are given in the Supplemental Experimental Procedures.

Surface Plasmon Resonance Analysis

Surface plasmon resonance spectra were measured using Biacore X or Biacore 3000 (GE Healthcare). Purified PAR1b(39-364) was immobilized on a CM5 sensor chip (GE Healthcare) according to the manufacturer's protocol. Recombinant CagA proteins at indicated concentrations in HBS-EP solution (GE Healthcare) were reacted with the immobilized PAR1b(39-364) at a flow rate of 20 μl/min. Dissociation constants were calculated using Biaevaluation software (GE Healthcare).

Static Light Scattering Analysis

Molecular weights (MWs) of recombinant CagA proteins were determined by the static light scattering (LS) method in a buffer of 20 mM Tris-HCl, 500 mM NaCl, pH 8.0. Sample solution was applied to the Superose 6 gel filtration column and the eluted sample was introduced to miniDAWN Tristar (Wyatt Technology) to measure the LS at 690 nm. The refractive index (RI) was also measured with RID-10A (Shimadzu). The MW was estimated from the LS/RI ratio.

Statistics

Statistics analysis was performed using either Student's t test or Mann-Whitney *U* test.

ACCESSION NUMBERS

Atomic coordinates and structure factors of CagA(1-876) and CagA(261-829) have been deposited in the Protein Data Bank under accession codes 4DVY and 4DVZ, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, two tables, three movies, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2012.05.010.

ACKNOWLEDGMENTS

We thank M. Matsuda and S. Hara for microinjection study and T. Shimizu, M. Nakamura, and D. Hishikawa for SPR analysis. We also thank S. Ohno for antibody and Y. Hirano and R. Natsume for help. GM/CA CAT is funded from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357. This work was supported by Grants-in-Aid for the Scientific Research on Innovative Area from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (M. Hatakeyama and T. Senda) and by the New Energy and Industrial Technology Development Organization (NEDO) of Japan (T. Senda).

Received: November 17, 2011
Revised: April 6, 2012
Accepted: May 7, 2012
Published: July 18, 2012

REFERENCES

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F_1 ATPase. Acta Crystallogr. D Biol. Crystallogr. *52*, 30–42.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

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

Backert, S., and Meyer, T.F. (2006). Type IV secretion systems and their effectors in bacterial pathogenesis. Curr. Opin. Microbiol. *9*, 207–217.

Bagnoli, F., Buti, L., Tompkins, L., Covacci, A., and Amieva, M.R. (2005). *Helicobacter pylori* CagA induces a transition from polarized to invasive phenotypes in MDCK cells. Proc. Natl. Acad. Sci. USA *102*, 16339–16344.

Blaser, M.J., and Atherton, J.C. (2004). *Helicobacter pylori* persistence: biology and disease. J. Clin. Invest. *113*, 321–333.

Buti, L., Spooner, E., Van den Veen, A., Rappuoli, R., Covacci, A., and Ploegh, H.L. (2011). Helicobacter pylori cytotoxin-associated gene A (CagA) subverts the apoptosis-stimulating protein of p53 (ASPP2) tumor suppressor pathway of the host. Proc. Natl. Acad. Sci. USA *108*, 9238–9243.

Covacci, A., and Rappuoli, R. (2000). Tyrosine-phosphorylated bacterial proteins: Trojan horses for the host cell. J. Exp. Med. *191*, 587–592.

Dunker, A.K., Silman, I., Uversky, V.N., and Sussman, J.L. (2008). Function and structure of inherently disordered proteins. Curr. Opin. Struct. Biol. *18*, 756–764.

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Fink, A.L. (2005). Natively unfolded proteins. Curr. Opin. Struct. Biol. *15*, 35–41.

Hatakeyama, M. (2004). Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. Nat. Rev. Cancer *4*, 688–694.

Hatakeyama, M. (2008). Linking epithelial polarity and carcinogenesis by multitasking *Helicobacter pylori* virulence factor CagA. Oncogene *27*, 7047–7054.

Higashi, H., Tsutsumi, R., Fujita, A., Yamazaki, S., Asaka, M., Azuma, T., and Hatakeyama, M. (2002a). Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. Proc. Natl. Acad. Sci. USA *99*, 14428–14433.

Higashi, H., Tsutsumi, R., Muto, S., Sugiyama, T., Azuma, T., Asaka, M., and Hatakeyama, M. (2002b). SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* Cag. Science *295*, 683–686.

Higashi, H., Yokoyama, K., Fujii, Y., Ren, S., Yuasa, H., Saadat, I., Murata-Kamiya, N., Azuma, T., and Hatakeyama, M. (2005). EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. J. Biol. Chem. *280*, 23130–23137.

Hohlfeld, S., Pattis, I., Püls, J., Plano, G.V., Haas, R., and Fischer, W. (2006). A C-terminal translocation signal is necessary, but not sufficient for type IV secretion of the *Helicobacter pylori* CagA protein. Mol. Microbiol. *59*, 1624–1637.

Holm, L., and Sander, C. (1996). Mapping the protein universe. Science *273*, 595–603.

Jiménez-Soto, L.F., Kutter, S., Sewald, X., Ertl, C., Weiss, E., Kapp, U., Rohde, M., Pirch, T., Jung, K., Retta, S.F., et al. (2009). *Helicobacter pylori* type IV secretion apparatus exploits β1 integrin in a novel RGD-independent manner. PLoS Pathog. *5*, e1000684.

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

Lu, H.S., Saito, Y., Umeda, M., Murata-Kamiya, N., Zhang, H.M., Higashi, H., and Hatakeyama, M. (2008). Structural and functional diversity in the PAR1b/MARK2-binding region of *Helicobacter pylori* CagA. Cancer Sci. *99*, 2004–2011.

Lu, H., Murata-Kamiya, N., Saito, Y., and Hatakeyama, M. (2009). Role of partitioning-defective 1/microtubule affinity-regulating kinases in the morphogenetic activity of *Helicobacter pylori* CagA. J. Biol. Chem. *284*, 23024–23036.

Mimuro, H., Suzuki, T., Tanaka, J., Asahi, M., Haas, R., and Sasakawa, C. (2002). Grb2 is a key mediator of *helicobacter pylori* CagA protein activities. Mol. Cell *10*, 745–755.

Moese, S., Selbach, M., Zimny-Arndt, U., Jungblut, P.R., Meyer, T.F., and Backert, S. (2001). Identification of a tyrosine-phosphorylated 35 kDa carboxy-terminal fragment (p35CagA) of the Helicobacter pylori CagA protein in phagocytic cells: processing or breakage? Proteomics *1*, 618–629.

Mohi, M.G., and Neel, B.G. (2007). The role of Shp2 (*PTPN11*) in cancer. Curr. Opin. Genet. Dev. *17*, 23–30.

Murata-Kamiya, N., Kikuchi, K., Hayashi, T., Higashi, H., and Hatakeyama, M. (2010). *Helicobacter pylori* exploits host membrane phosphatidylserine for delivery, localization, and pathophysiological action of the CagA oncoprotein. Cell Host Microbe *7*, 399–411.

Nagase, L., Murata-Kamiya, N., and Hatakeyama, M. (2011). Potentiation of *Helicobacter pylori* CagA protein virulence through homodimerization. J. Biol. Chem. *286*, 33622–33631.

Nesić, D., Miller, M.C., Quinkert, Z.T., Stein, M., Chait, B.T., and Stebbins, C.E. (2010). *Helicobacter pylori* CagA inhibits PAR1-MARK family kinases by mimicking host substrates. Nat. Struct. Mol. Biol. *17*, 130–132.

Ohnishi, N., Yuasa, H., Tanaka, S., Sawa, H., Miura, M., Matsui, A., Higashi, H., Musashi, M., Iwabuchi, K., Suzuki, M., et al. (2008). Transgenic expression of *Helicobacter pylori* CagA induces gastrointestinal and hematopoietic neoplasms in mouse. Proc. Natl. Acad. Sci. USA *105*, 1003–1008.

Parsonnet, J., Friedman, G.D., Orentreich, N., and Vogelman, H. (1997). Risk for gastric cancer in people with CagA positive or CagA negative *Helicobacter pylori* infection. Gut *40*, 297–301.

Pelz, C., Steininger, S., Weiss, C., Coscia, F., and Vogelmann, R. (2011). A novel inhibitory domain of *Helicobacter pylori* protein CagA reduces CagA effects on host cell biology. J. Biol. Chem. *286*, 8999–9008.

Ren, S., Higashi, H., Lu, H., Azuma, T., and Hatakeyama, M. (2006). Structural basis and functional consequence of *Helicobacter pylori* CagA multimerization in cells. J. Biol. Chem. *281*, 32344–32352.

Rohde, M., Püls, J., Buhrdorf, R., Fischer, W., and Haas, R. (2003). A novel sheathed surface organelle of the *Helicobacter pylori cag* type IV secretion system. Mol. Microbiol. *49*, 219–234.

Saadat, I., Higashi, H., Obuse, C., Umeda, M., Murata-Kamiya, N., Saito, Y., Lu, H., Ohnishi, N., Azuma, T., Suzuki, A., et al. (2007). *Helicobacter pylori* CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. Nature *447*, 330–333.

Segal, E.D., Cha, J., Lo, J., Falkow, S., and Tompkins, L.S. (1999). Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA *96*, 14559–14564.

Selbach, M., Paul, F.E., Brandt, S., Guye, P., Daumke, O., Backert, S., Dehio, C., and Mann, M. (2009). Host cell interactome of tyrosine-phosphorylated bacterial proteins. Cell Host Microbe *5*, 397–403.

Sigalov, A.B. (2010). Protein intrinsic disorder and oligomericity in cell signaling. Mol. Biosyst. *6*, 451–461.

Suzuki, M., Mimuro, H., Suzuki, T., Park, M., Yamamoto, T., and Sasakawa, C. (2005). Interaction of CagA with Crk plays an important role in *Helicobacter pylori*-induced loss of gastric epithelial cell adhesion. J. Exp. Med. *202*, 1235–1247.

Suzuki, M., Mimuro, H., Kiga, K., Fukumatsu, M., Ishijima, N., Morikawa, H., Nagai, S., Koyasu, S., Gilman, R.H., Kersulyte, D., et al. (2009). *Helicobacter pylori* CagA phosphorylation-independent function in epithelial proliferation and inflammation. Cell Host Microbe *5*, 23–34.

Tsang, Y.H., Lamb, A., Romero-Gallo, J., Huang, B., Ito, K., Peek, R.M., Jr., Ito, Y., and Chen, L.F. (2010). Helicobacter pylori CagA targets gastric tumor suppressor RUNX3 for proteasome-mediated degradation. Oncogene *29*, 5643–5650.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol. Biol. *364*, 215–230.

Xia, Y., Yamaoka, Y., Zhu, Q., Matha, I., and Gao, X. (2009). A comprehensive sequence and disease correlation analyses for the C-terminal region of CagA protein of *Helicobacter pylori*. PLoS ONE *4*, e7736.

Zeaiter, Z., Cohen, D., Müsch, A., Bagnoli, F., Covacci, A., and Stein, M. (2008). Analysis of detergent-resistant membranes of *Helicobacter pylori* infected gastric adenocarcinoma cells reveals a role for MARK2/Par1b in CagA-mediated disruption of cellular polarity. Cell. Microbiol. *10*, 781–794.