

# SCIENTIFIC REPORTS



OPEN

## Impact of structural polymorphism for the *Helicobacter pylori* CagA oncoprotein on binding to polarity-regulating kinase PAR1b

Received: 14 May 2016

Accepted: 28 June 2016

Published: 22 July 2016

Hiroko Nishikawa<sup>1,2,3</sup>, Takeru Hayashi<sup>1,2</sup>, Fumio Arisaka<sup>4</sup>, Toshiya Senda<sup>2,5</sup> & Masanori Hatakeyama<sup>1,2,3</sup>

Chronic infection with *cagA*-positive *Helicobacter pylori* is the strongest risk factor for atrophic gastritis, peptic ulcers, and gastric cancer. CagA, the product of the *cagA* gene, is a bacterial oncoprotein, which, upon delivery into gastric epithelial cells, binds to and inhibits the polarity-regulating kinase, partitioning-defective 1b (PAR1b) [also known as microtubule affinity-regulating kinase 2 (MARK2)], via its CagA multimerization (CM) motif. The inhibition of PAR1b elicits junctional and polarity defects, rendering cells susceptible to oncogenesis. Notably, the polymorphism in the CM motif has been identified among geographic variants of CagA, differing in either the copy number or the sequence composition. In this study, through quantitative analysis of the complex formation between CagA and PAR1b, we found that several CagA species have acquired elevated PAR1b-binding activity via duplication of the CM motifs, while others have lost their PAR1b-binding activity. We also found that strength of CagA-PAR1b interaction was proportional to the degrees of stress fiber formation and tight junctional disruption by CagA in gastric epithelial cells. These results indicate that the CM polymorphism is a determinant for the magnitude of CagA-mediated deregulation of the cytoskeletal system and thereby possibly affects disease outcome of *cagA*-positive *H. pylori* infection, including gastric cancer.

*Helicobacter pylori* is a spiral-shaped Gram-negative bacterium that colonizes approximately 50% of the world's population and is the causative agent of gastrointestinal diseases such as atrophic gastritis and peptic ulcers<sup>1,2</sup>. Infection with *H. pylori* is the strongest risk factor for the development of gastric cancer<sup>3,4</sup>, the third leading cause of cancer deaths worldwide<sup>5</sup>. Manifestations associated with chronic *H. pylori* infection vary considerably among distinct geographic regions and these differences have been attributed at least in part to polymorphisms of the *H. pylori* virulence factors such as CagA, VacA, IceA, BabA, DupA, and OipA<sup>6,7</sup>. Among those factors, much attention has been given to the structure-function relationship of CagA because of its strong association with gastric cancer<sup>8,9</sup>.

CagA, a 120~145-kDa protein, is injected into attached gastric epithelial cells via the bacterial type IV secretion system<sup>10,11</sup>. CagA comprises a structured N-terminal domain that tethers the protein to the inner leaflet of the plasma membrane<sup>12,13</sup> and an intrinsically disordered C-terminal tail that contains variable repeats of a tyrosine phosphorylation motif called the Glu-Pro-Ile-Tyr-Ala (EPIYA) motif<sup>14,15</sup>. Upon delivery into the host cell, the EPIYA motifs are tyrosine-phosphorylated by host kinases and thereby bind promiscuously to a number of SH2 domain-containing proteins, such as SHP2 tyrosine phosphatase<sup>16</sup>, C-terminal Src kinase (Csk)<sup>17</sup> and adaptor protein Crk<sup>18</sup>, to perturb their activities. Based on the amino acid sequences flanking these EPIYA motifs, four distinct EPIYA segments, EPIYA-A, -B, -C, and -D, with each segment carrying a single EPIYA motif, have been defined<sup>19</sup>. Approximately 60% of the *H. pylori* strains circulating all over the world, except East Asian countries such as Japan, China and Korea, produce Western CagA, a CagA species containing EPIYA-A, EPIYA-B and

<sup>1</sup>Division of Microbiology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan. <sup>2</sup>CREST, Japan Science and Technology Agency, Saitama 332-0012, Japan. <sup>3</sup>Max Planck-The University of Tokyo Center for Integrative Inflammation, Tokyo 113-0033, Japan. <sup>4</sup>College of Bioresource Sciences, Nihon University, Fujisawa 252-0880, Japan. <sup>5</sup>Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan. Correspondence and requests for materials should be addressed to M.H. (email: mhata@m.u-tokyo.ac.jp)

a variable number of EPIYA-C segments, typically present in 1–3 repeats in tandem<sup>2,19,20</sup>. On the other hand, almost all of the *H. pylori* strains isolated in East Asian countries produce a different class of CagA, termed the East Asian CagA, which, like Western CagA, contains the EPIYA-A and EPIYA-B segments but, instead of the EPIYA-C segments, carries a distinct segment termed EPIYA-D<sup>2,19</sup>. East Asian CagA has been linked to the high incidence of gastric cancer in East Asian countries compared to that in the rest of the world<sup>21</sup>. Among Western CagA species, those containing multiple EPIYA-C segments have been more closely associated with gastric cancer than those containing a single EPIYA-C segment<sup>20</sup>. In addition to the EPIYA tyrosine phosphorylation motif, the EPIYA-C segment also carries a 16-amino-acid sequence termed the CagA multimerization (CM) motif<sup>22</sup>. Initially identified as a motif that mediates CagA multimerization (dimerization), it was later discovered that the motif serves as a binding site for the polarity-regulating kinase, partitioning-defective 1 (PAR1), which can exist as a multimer (most probably dimer), in a phosphorylation-independent manner<sup>23</sup>. The crystal structure of the complex between PAR1b (39–364) and a fragment of Western CagA containing the CM motifs has revealed that the CM motif mimics host substrates to bind to the active site of PAR1b, inhibiting its kinase activity<sup>24</sup>. Like EPIYA segments, polymorphism also exists in the CM motif between Western CagA and East Asian CagA<sup>22</sup>. Western CagA typically possesses 2–4 repeats of the Western CM (CM<sup>W</sup>) motif as it usually contains 1–3 tandem repeats of the EPIYA-C segment that is followed by a single CM motif immediately downstream of this repeat region. East Asian CagA, in contrast, does not contain a CM motif in its EPIYA-D segment but carries a single East Asian CM (CM<sup>E</sup>) motif, which differs from the CM<sup>W</sup> by 5 amino acid residues, immediately downstream of the EPIYA-D segment<sup>22</sup>.

PAR1 is a highly conserved family of serine/threonine kinases, originally isolated in a screen for partitioning-defective mutants of *C. elegans* early embryos<sup>25</sup>. Independently, mammalian PAR1 was identified as microtubule affinity-regulating kinase, MARK, a family of kinases that phosphorylate microtubule-associated proteins, MAPs, such as tau, MAP2, and MAP4 at their microtubule-binding C-terminal repeats to release the MAPs from microtubules and destabilize them<sup>26,27</sup>. Mammalian PAR1 is comprised of four isoforms, PAR1a–d. The major isoform in epithelial cells, PAR1b, has been shown to localize at the basolateral membrane of polarized epithelial monolayers where it plays crucial roles in the establishment and maintenance of apical-basal epithelial cell polarity<sup>28–30</sup>. Accordingly, inhibition of PAR1b by *H. pylori* CagA in polarized epithelial monolayers provokes cell polarity and junctional defects<sup>23,31</sup>. Moreover, recent studies have indicated that the role of PAR1b in the cytoskeleton is not confined just to microtubule dynamics but also extends to filamentous actin (F-actin) rearrangements<sup>32,33</sup>. PAR1b inhibits a RhoA-specific GEF, GEF-H1, by phosphorylation to down-regulate RhoA and thereby inhibit stress fiber formation<sup>34</sup>. Stress fibers are contractile bundles of actomyosin anchored to focal adhesions, which generate the contractile force required for cell motility<sup>35,36</sup>.

Non-canonical CM motifs have been identified from other parts of the world such as Amerindian I CM (CM<sup>AmI</sup>) and Amerindian II CM (CM<sup>AmII</sup>) from an indigenous tribe of the Peruvian Amazon<sup>37–40</sup>. In this study, we examined for the first time the role of CM polymorphism in the CagA–PAR1b interaction through quantitative analysis and evaluated the impact of the CM polymorphism on the CagA action towards the actin cytoskeletal system to better understand the gastrointestinal pathogenesis caused by *cagA*-positive *H. pylori*.

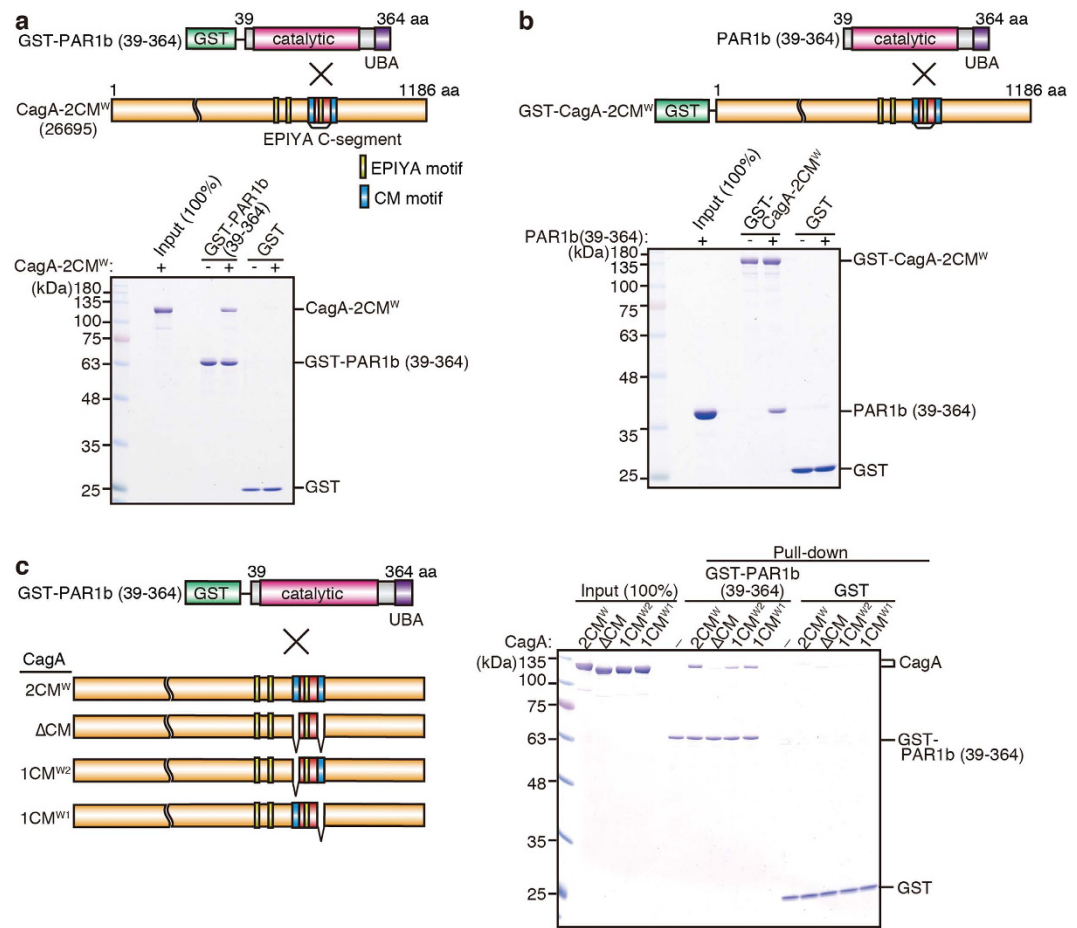
## Results

**CM motifs are required for CagA–PAR1b interaction *in vitro*.** Previous studies have shown that CagA and PAR1b specifically interact in COS-7 cells and human gastric AGS cells<sup>23,31,41</sup>. To determine whether the interaction is direct or not, recombinant PAR1b (39–364), which contains both the catalytic and the ubiquitin-associated (UBA) domains, was expressed as a GST fusion protein in *E. coli* and then purified (Fig. 1a, top). A canonical Western ABC type CagA (CagA-2CM<sup>W</sup>), which contains an EPIYA-C segment and thus two CM<sup>W</sup> motifs (one in the EPIYA-C segment and the other immediately downstream of EPIYA-C), was also expressed in *E. coli* and purified. A GST pull-down assay revealed that CagA-2CM<sup>W</sup> bound to GST-PAR1b (39–364), while it did not bind to the GST protein alone (Fig. 1a, bottom). To ascertain that the interaction is reproducible in a reciprocal pull-down, the GST-CagA-2CM<sup>W</sup> construct was used to pull-down PAR1b (39–364). As shown in Fig. 1b, GST-CagA-2CM<sup>W</sup> pulled down PAR1b (39–364), while the GST control protein did not (Fig. 1b, bottom). Collectively, these results indicated that CagA-2CM<sup>W</sup> interacted directly and specifically with PAR1b (39–364).

Next, to determine whether the CM motif is required for direct interaction of CagA with PAR1b *in vitro*, we generated three CagA-2CM<sup>W</sup> mutants: one that has the first CM (CagA-1CM<sup>W1</sup>), another that only has the second CM (CagA-1CM<sup>W2</sup>), and lastly, one that has no CM motif (CagA-ΔCM) (Fig. 1c, left). The recombinant proteins obtained were then mixed with GST-PAR1b (39–364) and a GST pull-down assay was performed. In accordance with our previous studies using cultured cells, GST-PAR1b (39–364) bound to CagA-2CM<sup>W</sup> but not to CagA-ΔCM<sup>23,31</sup> (Fig. 1c, right). The binding of GST-PAR1b (39–364) with CagA-2CM<sup>W</sup> was stronger than that with CagA-1CM<sup>W2</sup> or CagA-1CM<sup>W1</sup>, carrying only a single CM motif. We also observed a CM-dependent interaction between CagA and full-length PAR1b (Supplementary Fig. S1). Thus, at least a single CM motif was required for direct interaction of CagA with PAR1b.

## The two CM motifs in the C-terminus of CagA are not mutually exclusive in PAR1b binding.

Co-crystallization of the CagA peptide and PAR1b (39–364) has shown that the bound form of PAR1b (39–364) is approximately 40 Å in width<sup>24</sup>. Since the two CM motifs in CagA-2CM<sup>W</sup> reside in the intrinsically disordered region of CagA<sup>13</sup> and are 18 amino acid residues apart, or approximately 60 Å in the extended form, this region should be able to accommodate the binding of two PAR1b molecules (Fig. 2a). However, many intrinsically disordered proteins can induce folding upon binding to their target<sup>42</sup>. We therefore speculated that the binding of PAR1b (39–364) to one of the two CM motifs in CagA-2CM<sup>W</sup> might induce folding of the surrounding regions and inhibit the binding of PAR1b to the other CM motif. Thus, to determine whether the two CM motifs of CagA-2CM<sup>W</sup> can simultaneously bind to PAR1b or not, recombinant CagA-2CM<sup>W</sup> proteins with differential CM motif

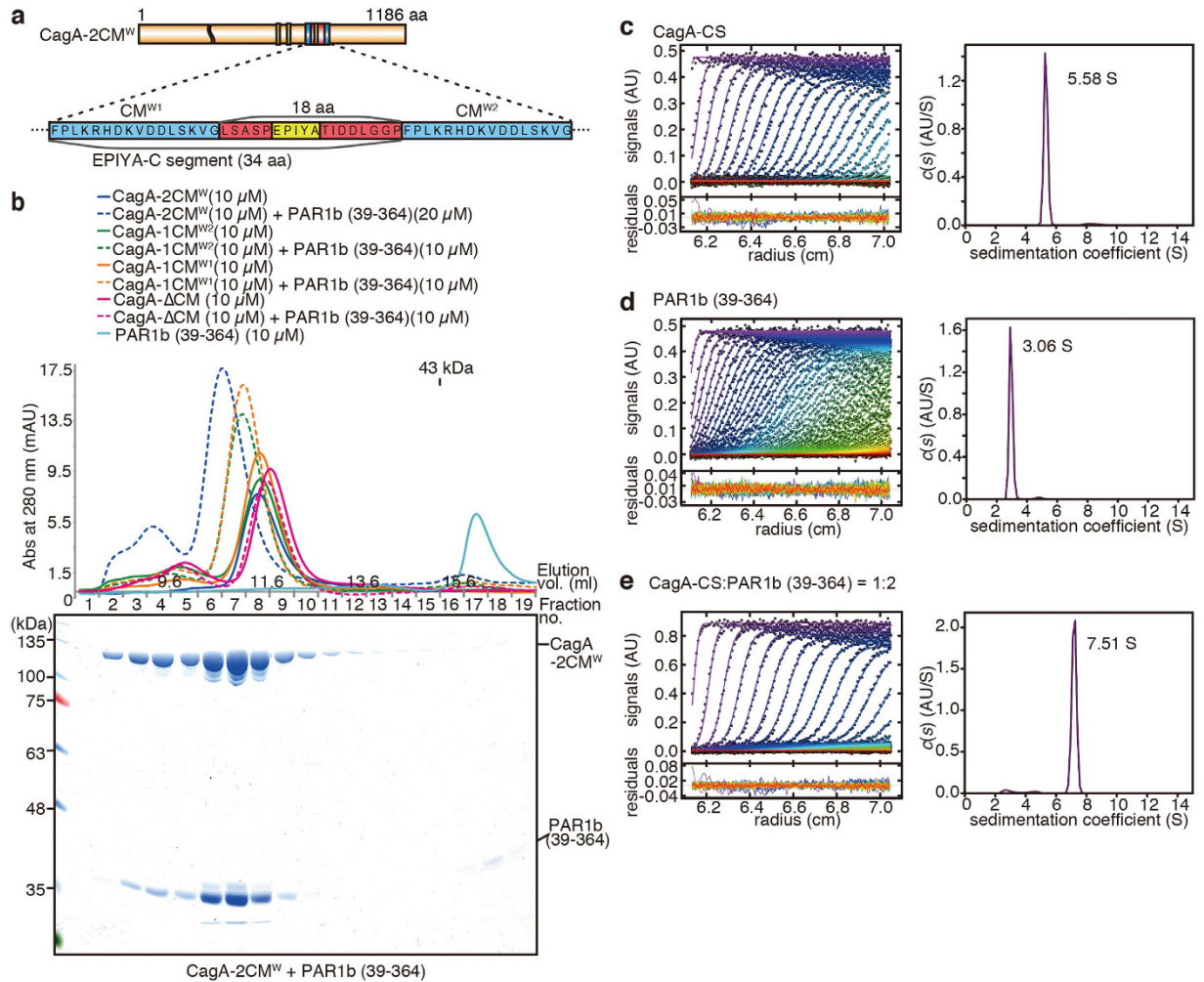


**Figure 1.** *In vitro* reconstitution of CagA-PAR1b interaction. **(a)** Schematic diagram of GST-PAR1b (39–364), which contains the catalytic and the UBA domains, and CagA-2CM<sup>W</sup> (*H. pylori* 26695 strain) which contains 2 CM motifs: one in the EPIYA-C segment and the other immediately downstream of the segment (*top*). Results for the GST pull-down between GST-PAR1b (39–364) and CagA-2CM<sup>W</sup> (*bottom*). **(b)** Schematic diagram of PAR1b (39–364) and GST-CagA-2CM<sup>W</sup> (*top*) and the results of the GST pull-down using these constructs (*bottom*). **(c)** Schematic diagram of GST-PAR1b (39–364) and GST-CagA-2CM<sup>W</sup> constructs with CM deletions (*left*) and the results of the GST pull-down using these constructs (*right*). All pull-downs were performed in HBS-P buffer. Samples were resolved on SDS-PAGE gels and visualized by CBB.

mutations were mixed *in vitro* with recombinant PAR1b (39–364) and analysed by size-exclusion chromatography (SEC) (Fig. 2b, top). When CagA-ΔCM, CagA-1CM<sup>W2</sup>, CagA-1CM<sup>W1</sup>, and CagA-2CM<sup>W</sup> were individually analyzed by SEC, the major peaks were very close to each other, eluting at 11.6 ml, 11.4 ml, 11.4 ml, and 11.4 ml, respectively. It should be noted that these major peaks represent monomeric CagA and the sub-peaks that elute earlier are due to artificial dimerization of CagA, caused *in vitro* via a C-terminal cysteine residue, which can oxidize to form disulphide bonds<sup>13</sup>. On the other hand, the major peak of PAR1b (39–364) was at 15.9 ml and was smaller than the 43-kDa marker. When CagA-ΔCM was mixed with PAR1b (39–364) at a molar ratio of 1:1, the major peak eluted in a position similar to the peak of CagA-ΔCM alone at 11.5 ml, suggesting that it does not bind to PAR1b (39–364). When CagA-1CM<sup>W2</sup> or CagA-1CM<sup>W1</sup> was mixed with PAR1b (39–364), major peaks for the two complexes eluted significantly earlier than peaks of CagA-1CM<sup>W2</sup> and CagA-1CM<sup>W1</sup> alone, at 11.0 ml. Notably, when CagA-2CM<sup>W</sup> and PAR1b (39–364) were mixed at a molar ratio of 1:2, the major peak eluted markedly earlier at 10.6 ml and the PAR1b (39–364) peak was significantly reduced, suggesting that most of PAR1b (39–364) formed complexes with CagA-2CM<sup>W</sup>. SEC fractions of CagA-2CM<sup>W</sup> and PAR1b (39–364) were then resolved on SDS-PAGE gel and stained by Coomassie Brilliant Blue (CBB) (Fig. 2b, bottom). Quantification of individual bands in fraction 7, the major peak, revealed that this peak contained CagA-2CM<sup>W</sup> and PAR1b (39–364) at a molar ratio of 1:2. From these observations, we concluded that the two CM motifs in a single CagA-2CM<sup>W</sup> molecule were not mutually exclusive in PAR1b binding, suggesting that two PAR1b molecules can simultaneously bind to a single CagA protein despite the proximity of the two CM motifs.

To further confirm that two PAR1b molecules can simultaneously bind to a single CagA-2CM<sup>W</sup> molecule, analytical ultracentrifugation (AUC) was performed. AUC has an advantage over SEC in that it provides a more precise measurement of molecular weights, which do not depend on the molecular shape. Indeed, SEC analysis previously showed that CagA-2CM<sup>W</sup>, whose sequence-based molecular weight is 134 kDa<sup>43</sup>, eluted earlier than

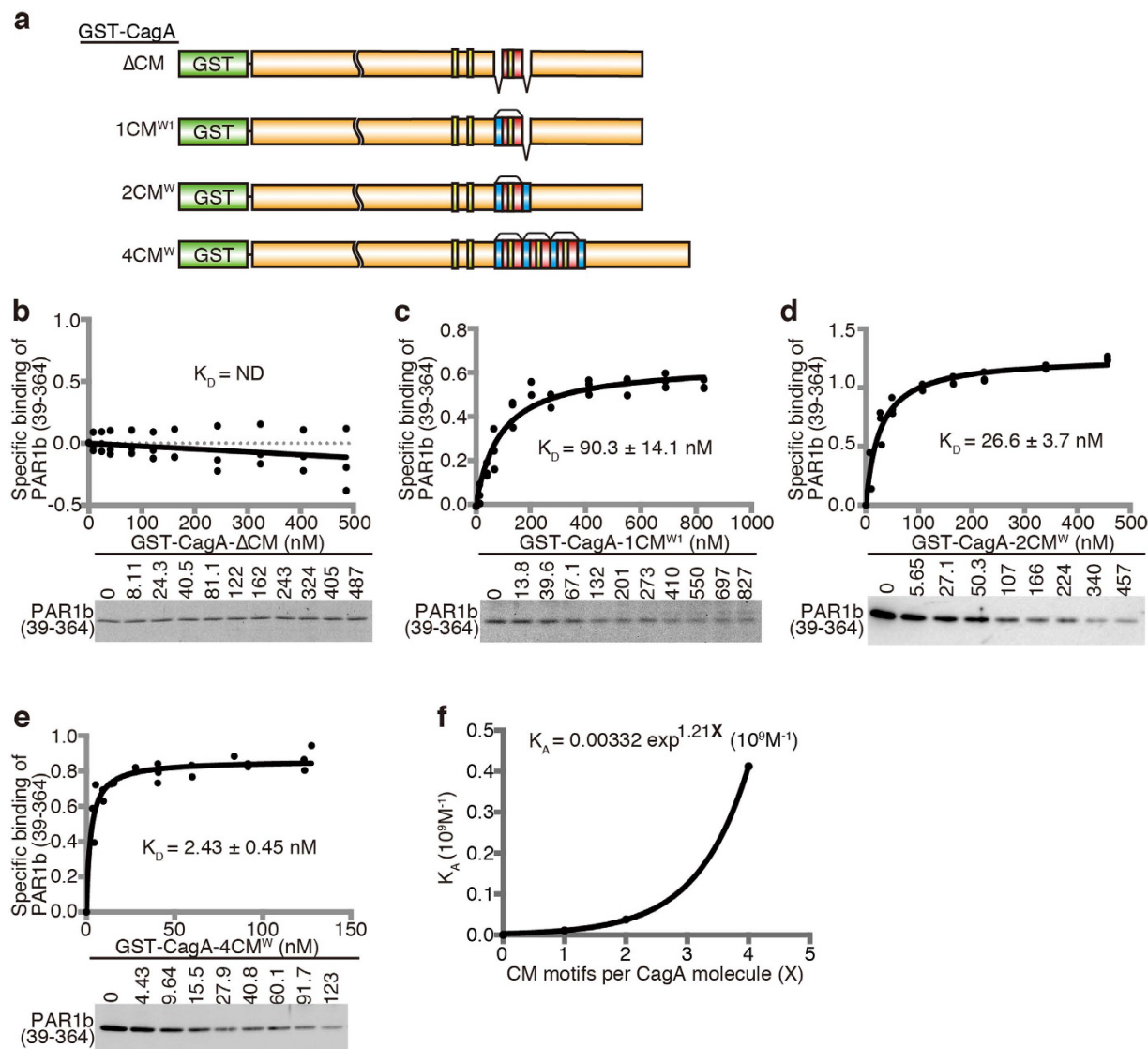




**Figure 2. Two PAR1b can simultaneously bind to CagA.** (a) Schematic diagram of CagA-2CM<sup>W</sup> and a blow-up of the region containing the CM motifs. CM<sup>W1</sup> resides in the 34-amino-acid EPIYA-C segment and CM<sup>W2</sup> flanks the segment on the C-terminal side. The two CM sequences are 18 amino acid residues apart. (b) Results from the size-exclusion chromatography of CagA-PAR1b (39–364) complexes analysed on Superdex 200 10/300 GL using HBS-P as running buffer (top). Fractionated CagA-2CM + PAR1b (39–364) complex was further resolved on SDS-PAGE gel and stained by CBB (bottom). (c–e) Analytical ultracentrifugation of 5.5 μM CagA-C1164S (CagA-CS) (c), 11 μM PAR1b (39–364) (d), and 5.5 μM CagA-CS + 11 μM PAR1b (39–364) (e). Raw absorbance distributions, the best-fit model (left), and the continuous sedimentation distributions  $c(s)$  (right) calculated by the SEDFIT program.

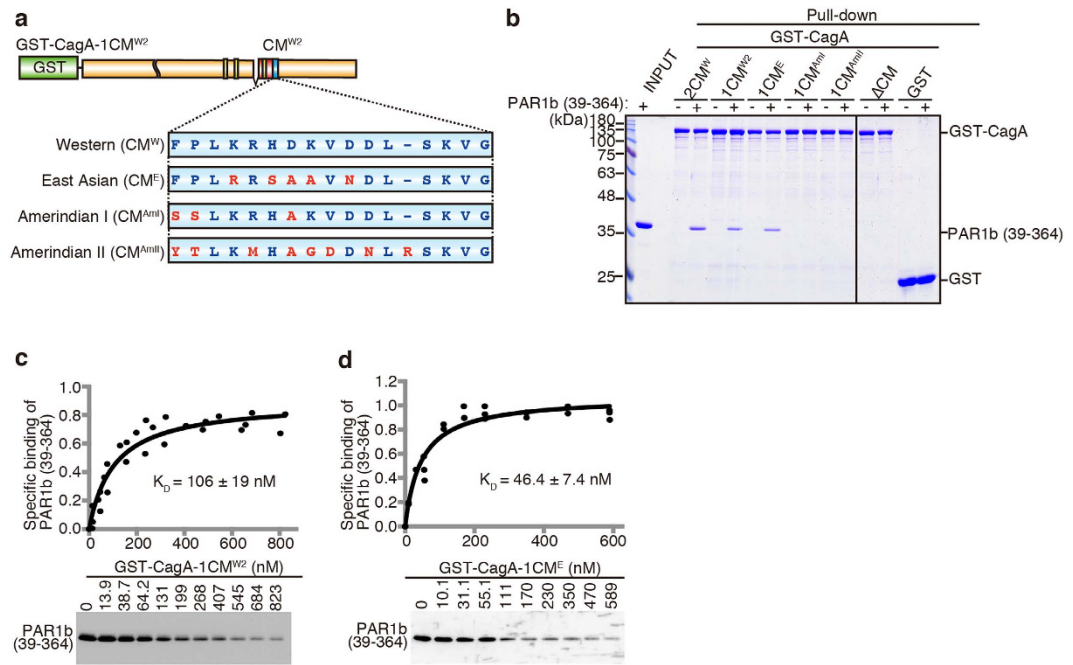
the 257-kDa marker, most likely due to its flat plate-like structure<sup>13</sup>. This time, to avoid artificial dimerization of CagA through oxidization, the oxidizable cysteine residue (C1164) of CagA-2CM<sup>W</sup> was mutated to a serine residue (CagA-CS). When CagA-CS and PAR1b (39–364) were analysed individually by AUC, they produced a peak with sedimentation coefficients of 5.58 S and 3.06 S, corresponding to the molecular weights of CagA-CS (121 kDa) and PAR1b (36.8 kDa), respectively (Fig. 2c,d). Since the sequence-based molecular weight of PAR1b (39–364) is 39.4 kDa, the AUC results are consistent with the conclusion that the two recombinant proteins existed as monomers in solution. When CagA-CS and PAR1b (39–364) were mixed at a ratio of 1:2 and analysed by AUC, it produced a sharp major peak with a sedimentation coefficient of 7.51 S (93% of the total loading concentration; Fig. 2e). Conversion of  $c(s)$  to  $c(M)$  gave the molecular weight of 182 kDa which was slightly (6%) smaller than that of the expected 1:2 complex but (15%) larger than that of the 1:1 complex. In addition, the remaining monomeric molecular species of CagA-CS and PAR1b (39–364) specimens in the solution were less than 2 or 3% of the loading concentration. We therefore concluded that CagA-CS and PAR1b interact in a 1:2 stoichiometry. Thus, the results from the AUC experiment were in agreement with the results from the SEC experiment and provided additional evidence that a CagA-2CM<sup>W</sup> molecule was capable of simultaneously interacting with two PAR1b molecules via duplicated CM motifs.

**Tandem CM motifs in Western CagA synergize in PAR1b binding.** Since clinically isolated Western CagA species generally carry 2–4 copies of the CM<sup>W</sup> motif, which are not mutually exclusive to each other in



**Figure 3. Tandem CM motifs generate strong binding affinity to PAR1b.** (a) Schematic diagram of GST-CagA constructs used. CagA-4CM<sup>W</sup> was derived from *H. pylori* NCTC11637 strain. (b–e) Saturation binding curves obtained by quantitative GST binding assay between PAR1b (39–364) and GST-CagA-ΔCM (b), GST-CagA-1CM<sup>W</sup> (c), GST-CagA-2CM<sup>W</sup> (d), or GST-CagA-4CM<sup>W</sup> (e). Values shown are dissociation constants ( $K_D$ ) ± SE, n = 3. Bands at the bottom of each graph shows a representative CBB stained gel or an immunoblot of unbound PAR1b (39–364). Full-length gels and blots are presented in Supplementary Fig. S2. (f) Graph and equation depicting the relationship between the number of CM motifs per molecule of CagA and their association constants ( $K_A$ ) with PAR1b (39–364).

terms of binding to PAR1b, we investigated whether carrying multiple repeats of CM<sup>W</sup> motifs in close proximity confers an advantage in binding to PAR1b by directly comparing the binding affinities of CagA constructs with different numbers of CM<sup>W</sup> motifs to PAR1b (39–364). To obtain quantitative measurements of the binding affinities between the CagA constructs and PAR1b (39–364), we used the quantitative GST pull-down assay<sup>44</sup>. To this end, a GST tag was fused to CagA containing 0–4 repeats of the CM<sup>W</sup> motifs (Fig. 3a). The resulting GST-CagA fusion proteins were bound to beads and aliquoted to each reaction tube in a wide range of concentrations and then mixed with a fixed low concentration of PAR1b (39–364). Unbound PAR1b (39–364) was quantified after running the supernatant of the reaction on an SDS-PAGE gel. Bound PAR1b (39–364) was calculated by subtracting the unbound PAR1b (39–364) from total PAR1b (39–364) and then a saturation binding curve was obtained to determine the apparent dissociation constant ( $K_D$ ) (Fig. 3b–e). The  $K_D$  value for CagA-ΔCM could not be determined as no significant binding was observed (Fig. 3b), in accordance with the results shown in Figs 1c and 2b. The binding affinity, as determined by the  $K_D$  value, strengthened in correlation with the number of tandem CM<sup>W</sup> motifs. CagA-1CM<sup>W</sup> and CagA-2CM<sup>W</sup> exhibited  $K_D$  values of  $90.3 \pm 14.1 \text{ nM}$  and  $26.6 \pm 3.7 \text{ nM}$ , respectively (Fig. 3c,d). Strikingly, CagA-4CM<sup>W</sup>, containing four CM<sup>W</sup> motifs, displayed a marked increase in



**Figure 4. Binding affinities of regional CagA CM variations to PAR1b.** (a) Schematic diagram of GST-CagA-1CM<sup>W2</sup> carrying various regional CM motifs at the CM<sup>W2</sup> position. (b) GST pull-down between various GST-CagA-1CM<sup>X</sup> constructs with various CM motifs and PAR1b (39–364). (c,d) Saturation binding curves obtained by the quantitative GST binding assay between PAR1b (39–364) and GST-CagA-1CM<sup>W2</sup> (c) or GST-CagA-1CM<sup>E</sup> (d). Values are dissociation constants ( $K_D$ )  $\pm$  SE,  $n = 3$ . Full-length gels and blots are presented in Supplementary Fig. S3.

PAR1b-binding affinity ( $K_D = 2.43 \pm 0.45$  nM), which was more than 10-fold greater than that of CagA-2CM<sup>W</sup> and more than 37-fold greater than that of CagA-1CM<sup>W1</sup>.

To further investigate the relationship between the number of CM motifs per CagA molecule and the binding affinity of the CagA-PAR1b interaction, the apparent association constant ( $K_A = 1/K_D$ ) was plotted against the number of tandem CM motifs (Fig. 3f). If the CM motifs bind to PAR1b (39–364) independently of each other and the  $K_A$  value of CagA-PAR1b increases only as a result of the increase in the concentration of CM motifs, then the  $K_A$  value should increase proportionally to the number of CM motifs per CagA molecule. However, this was clearly not the case; an exponential curve of  $K_A = 0.00332 \exp^{1.21X} 10^9 \text{M}^{-1}$  (where X is the number of CM motifs) was obtained by curve fitting. The result therefore indicated that the tandem CM motifs in a CagA molecule act synergistically in generating strong PAR1b binding, which would otherwise require a very high concentration of intracellular CagA if only a single CM motif existed per CagA molecule.

**Binding affinities of geographic CM variations to PAR1b.** While Western CagA species are known to contain multiple (usually 2–4) copies of the CM<sup>W</sup> motif, East Asian CagA species contain only a single copy of the CM<sup>E</sup> motif, which differs by 5 amino acids from its Western counterpart<sup>22</sup>. Interestingly, this sequence alteration seems to be crucial for the CM<sup>E</sup> motif in generating a stronger binding affinity to PAR1b than the CM<sup>W</sup> motif<sup>31</sup>. Additional variations in the CM motif have also been discovered from other parts of the world<sup>37,38,40</sup>. They include CM<sup>AmI</sup> and CM<sup>AmII</sup> motifs from Amerindian natives of the Peruvian Amazon<sup>37,39</sup>.

To compare the PAR1b binding activities of these geographic CM variants with the binding activity of CM<sup>W</sup>, we made use of the approach used for determining PAR1b binding of CagA-2CM<sup>W</sup> (Fig. 1c). Since CM<sup>E</sup> and CM<sup>AmII</sup> motifs usually exist as a single copy at the CM<sup>W2</sup> position of their respective geographic CagA variants, we generated chimeric GST-CagA constructs by replacing the CM<sup>W2</sup> motif of GST-CagA-1CM<sup>W2</sup> with that of the various geographic CM variants (GST-CagA-1CM<sup>X</sup>; X as the variant) (Fig. 4a). A GST pull-down assay showed that GST-CagA-1CM<sup>E</sup> bound to PAR1b (39–364) as previously described<sup>31</sup> (Fig. 4b). In contrast, neither GST-CagA-1CM<sup>AmI</sup> nor GST-CagA-1CM<sup>AmII</sup> bound to PAR1b (39–364).

For a more quantitative evaluation of the binding affinities of CM<sup>W</sup> and CM<sup>E</sup> motifs to PAR1b, we performed quantitative GST pull-down assays using the chimeric GST-CagA proteins to determine the  $K_D$  values. As shown in Fig. 4c, the  $K_D$  value between GST-CagA-1CM<sup>W2</sup>, which contains a single CM<sup>W</sup> motif, and PAR1b (39–364) was  $106 \pm 19$  nM, whereas the  $K_D$  value between GST-CagA-1CM<sup>E</sup>, which contains a single CM<sup>E</sup> motif, and PAR1b (39–364) was  $46.4 \pm 7.4$  nM. Hence, the binding affinity of the CM<sup>E</sup> motif to PAR1b (39–364) was twice as strong as that of the CM<sup>W</sup> motif, being consistent with the results obtained from co-immunoprecipitation assays using a COS-7 overexpression system which showed that the binding activity of a single CM<sup>E</sup> motif is equivalent to that of two tandem CM<sup>W</sup> motifs<sup>31</sup>.  $K_D$  values determined in this study are summarized in Table 1. It showed that the binding affinity of the CM<sup>W</sup> motif to PAR1b was constant irrespective of its position at CM<sup>W1</sup> or CM<sup>W2</sup>



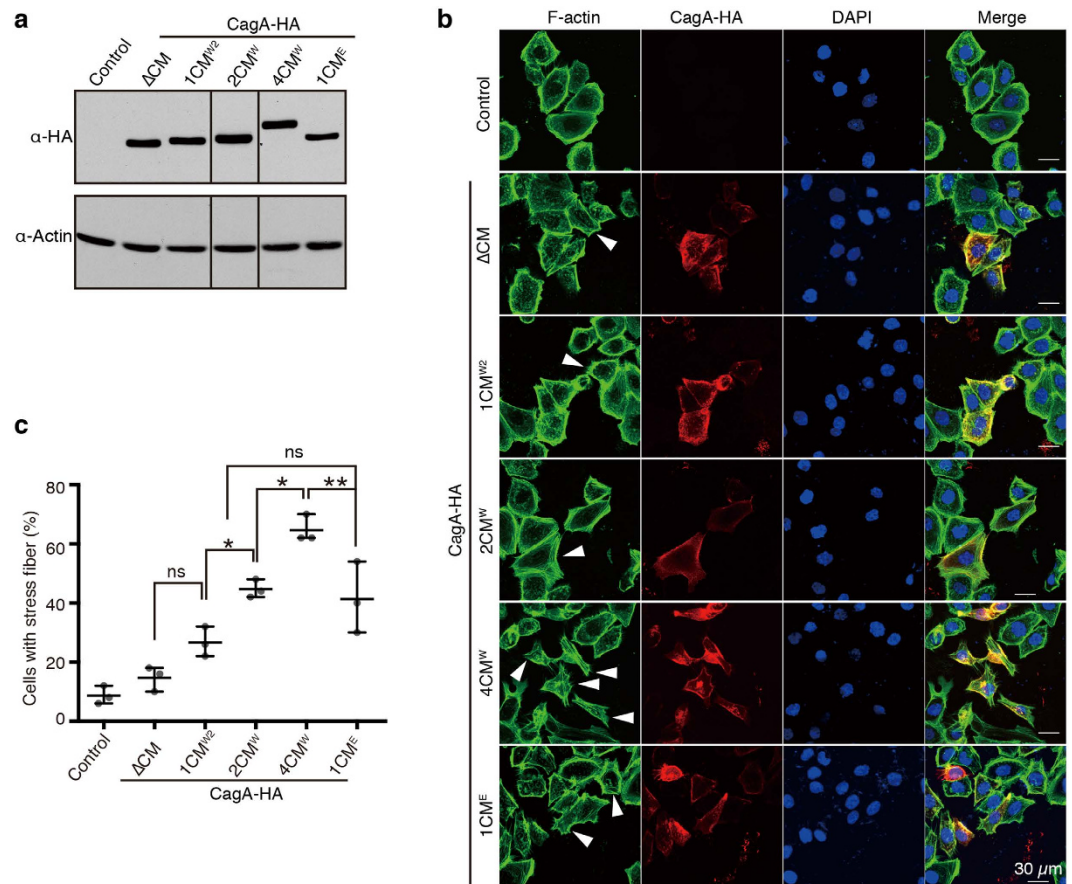
GST-CagA constructs	Type of CM motif	CM sequence	No. of CM motifs	PAR1b binding	$K_D \pm SE$ (nM)
$\Delta$ CM	—	—	0	N	ND
1CM <sup>W1</sup>	Western	FPLKRHDKVDDLKSKVG	1	Y	90.3 $\pm$ 14.1
2CM <sup>W</sup>	Western	FPLKRHDKVDDLKSKVG	2	Y	26.6 $\pm$ 3.7
4CM <sup>W</sup>	Western	FPLKRHDKVDDLKSKVG	4	Y	2.43 $\pm$ 0.45
1CM <sup>W2</sup>	Western	FPLKRHDKVDDLKSKVG	1	Y	106 $\pm$ 19
1CM <sup>E</sup>	East Asian	FPLRRSAAVNDLKSKVG	1	Y	46.4 $\pm$ 7.4
1CM <sup>AmlI</sup>	Amerindian I	SSLKRHAKVDDLKSKVG	1	N	ND
1CM <sup>AmlII</sup>	Amerindian II	YTLKMHAGDDNLRKSKVG	1	N	ND

**Table 1. Summary table of binding affinities between various CagA constructs and PAR1b.** ND: not determined.

in CagA. Table 1 also showed that while the binding affinity of CagA containing a single CM<sup>E</sup> motif was twice as strong as that of CagA containing a single CM<sup>W</sup> motif, duplication of the CM<sup>W</sup> motif in the context of CagA potentiated the PAR1b-binding affinity of CagA to a level comparable to that of CagA possessing a single CM<sup>E</sup> motif.

**Augmentation of CagA-PAR1b binding potentiates stress fiber formation in cells.** Western CagA species was capable of undergoing strong PAR1b binding through tandem duplication of the CM<sup>W</sup> motif. Also, East Asian CagA carrying a single CM<sup>E</sup> motif exhibited a PAR1b-binding affinity similar to that of Western CagA carrying two CM<sup>W</sup> motifs *in vitro*. However, it was possible that, in cells, CagA with a single CM<sup>W</sup> motif would be sufficient in saturating endogenous PAR1b binding. Moreover, the EPIYA-C motif, which serves as an SHP2-binding site upon tyrosine phosphorylation, exists in the 18-amino-acid spacer between the tandem CM<sup>W</sup> motifs, and the CagA-SHP2 interaction could competitively inhibit CagA-PAR1b interaction in a cellular context. Given this, we decided to assess the biological impact of different CM polymorphisms with regard to the cytoskeletal role of PAR1b. To this end, we examined the effect of CagA on the formation of stress fibers in AGS human gastric epithelial cells. Stress fibers are contractile bundles of actomyosin that play an important role in cell adhesion, cell motility, and morphogenesis<sup>35,36</sup>. We previously showed that PAR1b regulates stress fiber formation by suppressing RhoA activity via inhibition of a Rho-specific GEF, GEF-H1<sup>34</sup>. If CagA with more CM motifs inhibits endogenous PAR1b more efficiently than does CagA with fewer CM motifs through enhanced complex formation, then cultured cells expressing CagA species with more CM motifs should display greater stress fiber formation. Thus, HA-tagged constructs of CagA- $\Delta$ CM, CagA-1CM<sup>W2</sup>, CagA-2CM<sup>W</sup>, CagA-4CM<sup>W</sup>, and CagA-1CM<sup>E</sup> were transiently expressed in AGS cells. The cells were then immunostained with an anti-HA antibody to identify CagA-HA-expressing cells (Fig. 5a,b). F-actin was visualized by fluorescent-labelled phalloidin, and the magnitude of stress fiber formation was quantitatively assessed. CagA was localized to the membrane as previously described<sup>16,45</sup>. The majority of AGS cells expressing CagA- $\Delta$ CM displayed strong cortical staining of F-actin, similar to the control cells, while only 15  $\pm$  5% of CagA-positive cells exhibited stress fiber formation (Fig. 5c). Cells expressing CagA with a single copy of the CM motif, CagA-1CM<sup>W2</sup>, did not show a significant increase ( $P = 0.157$ ) in the proportion of cells with stress fibers (27  $\pm$  5%) compared to cells expressing CagA- $\Delta$ CM (15  $\pm$  5%). However a significant increase in the proportion of cells displaying stress fiber formation was observed in cells expressing CagA-2CM<sup>W</sup> (45  $\pm$  3%;  $P = 0.020$  compared to CagA-1CM<sup>W2</sup>) and a further increase was observed in cells expressing CagA-4CM<sup>W</sup> (65  $\pm$  5%;  $P = 0.010$  compared to CagA-2CM<sup>W</sup>). Of note, the morphological change observed in AGS cells expressing CagA-4CM<sup>W</sup> was attributable to SHP2 deregulation by CagA, the activity of which is proportional to the number of EPIYA-C segments<sup>46</sup>. While we cannot exclude the potential contribution of EPIYA-C segments to stress fiber formation in cells expressing CagA-4CM<sup>W</sup>, the other three constructs, CagA- $\Delta$ CM, CagA-1CM<sup>W2</sup> and CagA-2CM<sup>W</sup>, all contain only a single EPIYA-C motif and thus can be compared directly. Tandem duplication of CM motifs in CagA molecules may therefore increase the proportion of cells displaying stress fiber formation, which reflects the degree of CagA binding to PAR1b in cells. Furthermore, when the chimeric CagA-1CM<sup>E</sup>, which carries a single CM<sup>E</sup> motif, was transiently expressed in AGS cells, the number of cells displaying stress fiber formation was similar to that of cells expressing CagA-2CM<sup>W</sup> (Fig. 5c;  $P = 0.970$ ), also being consistent with our *in vitro* results showing that CagA with two CM<sup>W</sup> motifs is similar to CagA with one CM<sup>E</sup> motif in terms of PAR1b binding (Table 1).

**Inverse correlation between the number of CagA CM motifs and transepithelial electrical resistance (TER).** Finally, we decided to investigate the effects of the series of CagA CM mutants on cell-cell interaction by measuring transepithelial electrical resistance (TER), which reflects the strength of the tight junction, using polarized Madin-Darby canine kidney (MDCK) epithelial monolayers. The experiment makes it possible to provide objective data for quantitatively evaluating the degree by which each CagA mutant perturbs tight junction integrity. Previous application of TER showed that, when expressed in polarized MDCK monolayers, CagA-1CM<sup>E</sup> disrupts the tight junction more than CagA-1CM<sup>W2,31</sup>. To address whether the disruption of the tight junctions is proportional to the number of CM<sup>W</sup> motifs, we infected polarized MDCK cells with lentivirus transducing copGFP (negative control), CagA-1CM<sup>W2</sup>, or CagA-2CM<sup>W</sup> and measured TER across the polarized MDCK monolayer for 3 days (Fig. 6a,b). Two days after transduction, CagA-2CM<sup>W</sup> elicited a sharp decrease in TER compared to control, and the effect seemed reach a saturation point at Day 3. CagA-1CM<sup>W2</sup>, on the other hand, disrupted the tight junctions more slowly than CagA-2CM<sup>W</sup>. Thus, the results of the experiment revealed



**Figure 5. CM-dependent augmentation of stress fiber formation by CagA.** (a) Immunoblots of AGS total cell lysates 24 h after transfection with the various pCDH-EF1-CagA-HA expression vectors as indicated. CagA-HA was detected by using anti-HA (3F10) as primary antibody. Control indicates cells transfected with the empty pCDH-EF1 vector. Full-length blot is presented in Supplementary Fig. S4. (b) Representative fluorescent images of AGS cells after transfection. F-actin was visualized by Alexa Fluor 488 phalloidin. CagA-HA was detected using anti-HA (C29F4) followed by Alexa Fluor 546 anti-rabbit antibodies. Nuclei were stained with DAPI. Arrowheads mark CagA-HA expressing cells positive for stress fiber. (c) Quantification of CagA-HA expressing AGS cells displaying stress fiber. 50 CagA-positive cells per construct were analysed for stress fiber formation per experiment. For control cells, 50% of cells in each image were chosen randomly and assessed for stress fiber formation. Error bars represent mean  $\pm$  range, n = 3. \* $P$  < 0.05, \*\* $P$  < 0.01, repeated measures one-way ANOVA, post-hoc Tukey's test.

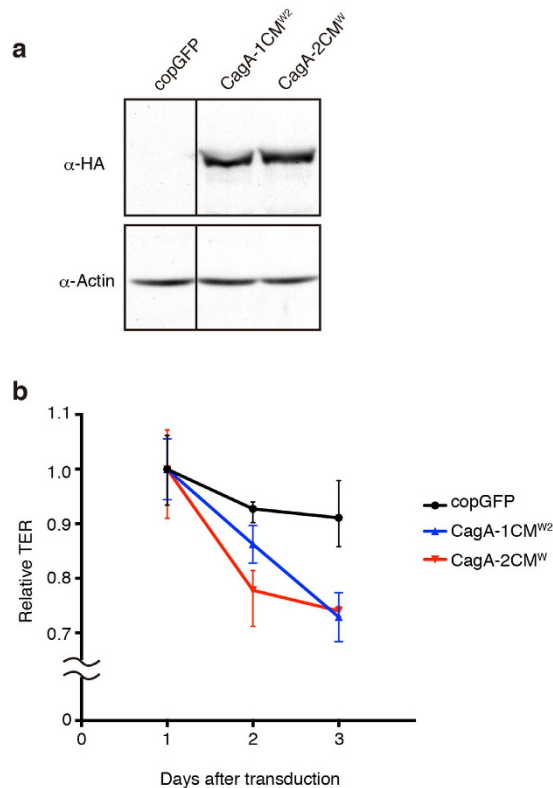
that the degree of CagA-mediated disruption of tight junctions was proportional to the number of CM motifs, which determines the binding strength of CagA to PAR1b and subsequent inhibition of PAR1b kinase activity.

## Discussion

Structural polymorphisms in the C-terminal tail of CagA, which correlate with the differential geographical distribution of *H. pylori*, have been implicated to influence the development of gastrointestinal diseases<sup>8,9</sup>. In this study, we investigated the CM motif by using quantitative methods to address how the structural polymorphism of this motif influences the PAR1b-binding activity of CagA. We found that in the case of Western CagA, an increase in the number of CM motifs in a CagA molecule augments its binding affinity to PAR1b in an exponential manner (Fig. 3f). Strikingly, CagA-4CM<sup>W</sup> generated a  $K_D$  value that is more than 37-fold higher than that of CagA-1CM<sup>W1</sup>. This robust augmentation of the CagA-PAR1b interaction can be explained by the fact that when the first CM motif of a CagA molecule interacts with a PAR1b molecule, it brings the other CM motifs to the vicinity of the PAR1b molecule. Upon dissociation of PAR1b from the first CM motif, the other CM motifs can immediately recapture PAR1b. Thus, an increase in the CM copy number would increase the chances of recapturing PAR1b on the same CagA molecule, potentiating the overall binding affinity of the CagA-PAR1b interaction.

Intriguingly, a single CM<sup>E</sup> motif can generate a binding affinity to PAR1b that is similar to that of two tandem CM<sup>W</sup> motifs (Fig. 3d and 4d). Hence *H. pylori* has evolved two different strategies for efficiently inhibiting endogenous PAR1b. The first strategy is quantitative, by increasing the number of tandem CM motifs, as is the case with Western CagA, and the second strategy is qualitative, by base substitution within the CM motif, as is the case with East Asian CagA. We also showed that neither CM<sup>AmlI</sup> nor CM<sup>AmlII</sup> motifs bound to PAR1b, confirming the





**Figure 6. CM-dependent disruption of polarized MDCK monolayers.** (a) Immunoblots of MDCK total cell lysates 3 days after transduction with the lentiviral constructs as indicated. CagA-HA was detected by using anti-HA (6E2) as primary antibody. The copGFP-transducing lentivirus was used as a negative control. Full-length blot is presented in Supplementary Fig. S5. (b) Relative TER of polarized MDCK monolayers normalized to Day 1 post transduction. Error bars represent mean  $\pm$  range, n = 3.

results of a previous co-immunoprecipitation study showing that the binding of PAR1b to the Amerindian CagA proteins is extremely weak, if any<sup>39</sup>.

PAR1b is a master regulator for the establishment and maintenance of polarity<sup>30</sup>. Inhibition of this kinase by CagA has been shown to cause disruption of the tight junctions, loss of polarity<sup>23</sup>, microtubule disruption<sup>32</sup>, and promotion of transcytosis<sup>47</sup>. The dominant prevalence of East Asian CagA and Western CagA subtypes in much of the world suggests that either a single copy of CM<sup>E</sup> or two CM<sup>W</sup> motifs enables optimal inhibition of PAR1b, which contributes to successful colonization of *H. pylori*. *H. pylori* lives in a surprisingly nutrient-deprived environment, and iron uptake from the host has been shown to be a critical factor for the success of *H. pylori* colonization, as iron in the host's stomach is not in its usable form<sup>47,48</sup>. The bacterium therefore disrupts host cell polarity, most likely through PAR1b inhibition, to promote transcytosis of iron from the basolateral side of the host cell<sup>47</sup>. Interestingly, *H. pylori* harvested from an iron-depleted Mongolian gerbil model has been reported to exhibit enhanced virulence and to deliver elevated levels of CagA into host cells, presumably to obtain iron by promoting more transcytosis<sup>49</sup>. While in normal circumstances, Western *H. pylori* expressing CagA-ABC, which contains two CM<sup>W</sup> sequences, may prevail in the stomach due to higher resistance to low pH<sup>20</sup>, selective pressures in a more nutrient-deprived environment may give substantial advantage to *H. pylori* strains that express CagA with larger numbers of CM motifs, in a mixed population. In such a case, if the C-terminal tail were structured, the protein would have to be larger to accommodate for multiple binding sites, making it more costly and disadvantageous for *H. pylori* to produce CagA with limited resources<sup>50,51</sup>. Thus, from *H. pylori*'s point of view, duplication of CM motifs in an intrinsically disordered region may offer a cost-effective solution to produce a potent effector protein. On a similar note, the Amerindian *H. pylori* strains have been outcompeted considerably by Western strains in the Latin American Mestizo and urban Amerindian populations, which has been attributed to their reduced fitness compared to Western strains<sup>37–39,52</sup>. Part of this reduced fitness may arise from the inability of Amerindian CagA to bind PAR1b as shown in this study.

The evolution of virulence factors in pathogens is influenced by a balance between the costs and benefits of virulence. Depending on the size of host populations and the degree of pathological impacts, pathogens may evolve toward either increasing or decreasing virulence because the death or serious disability of the host substantially influences the expansion of pathogens. Loss of PAR1b-binding activity in CagA indicates that such a CagA variant only weakly induces mucosal damage and thereby ameliorates immune/inflammatory responses, which enable long-term *H. pylori* colonization. At the same time, however, reduced virulence restrains infectivity of *H. pylori* and subsequent disease development. Thus, only in a particular society such as small and extremely

isolated Amazonian rainforest tribes<sup>37,38,40</sup>, attenuation of CagA virulence could become an affordable option in preventing *H. pylori* from extinction.

The *in vitro* mixing experiment carried out in this study revealed successful interaction of a CagA monomer (ABC-type Western CagA possessing two CM motifs) with two PAR1 molecules simultaneously. We previously reported that CagA can dimerize in the host cells it has been delivered to<sup>22,53</sup>. Mechanistically, this CagA dimerization is mediated by the interaction of CagA with PAR1b, which is present as a dimer, via the CM motif<sup>23,24,31,53</sup>. Since the recombinant PAR1b fragment (residues 39–364 of full-length PAR1b, which comprises 788 amino acid residues) used in this study was a monomer in solution, PAR1b dimerization may be mediated via the N-terminal or C-terminal region, which is lacking in the PAR1b fragment. Alternatively, dimerization of PAR1b might require post-translational modification or another cellular protein(s) in the mammalian host cells. In any case, our results indicate that a single CagA containing two or more CM motifs can accommodate multiple PAR1 molecules, either monomeric or dimeric forms, and thereby collectively inhibits their kinase activity.

Here, we have shown for the first time that CagA can induce stress fiber formation in a CM-dependent manner presumably through activating RhoA. Furthermore, the degree of stress fiber formation in cultured cells was in accordance with the ability of the CM polymorphism to bind PAR1b *in vitro*. Formation of stress fibers involves remodeling of actin cortical bundles into contractile actin fibers, and anchoring of these contractile fibers to focal adhesion for mechanotransduction, which may facilitate cell migration. Moreover, we showed that when CagA is transduced in polarized MDCK monolayers, the number of CM affects the acuteness in which the tight junction is disrupted. Therefore, in conjunction with the loss of tight junctions and cell polarity, stress fibers may contribute to oncogenesis by promoting invasion of surrounding tissue and metastatic dissemination<sup>54</sup>. It has been reported that expression of constitutively active RhoA in Madin-Darby kidney (MDCK) cells enhances stress fiber formation and causes cell morphological changes, resulting in gaps in a monolayer<sup>55</sup>. Thus, strong inhibition of PAR1b by CagA may compromise the integrity of the gastric epithelial layer via two mechanisms: (1) disruption of tight junctions by causing polarity defects and (2) actin remodeling via RhoA activation, thereby resulting in severe gastritis and peptic ulcers<sup>23,34</sup>. A recent study has shown that tandem duplication of the EPIYA-C segment robustly enhances SHP2-binding affinity of CagA<sup>46</sup>. Accordingly, strengthened binding activities to both PAR1b and SHP2 conceivably contribute to the greater incidence of gastric cancer associated with infection with *H. pylori* strains producing CagA with multiple EPIYA-C segments.

In conclusion, the degree of pathophysiological actions of *H. pylori* CagA is influenced by the sequence polymorphism of the CagA CM motif. The CM diversity may therefore contribute to the different disease manifestations caused by *cagA*-positive *H. pylori* infection. The present work will help in understanding the molecular mechanisms underlying clinical outcomes of *H. pylori* infection. Its multifaceted functions make PAR1b an attractive target for *H. pylori* in manipulating the cell host for successful colonization of the stomach, while provoking adverse and pathological effects for the host in the long term.

## Methods

**Bacterial expression vectors.** pGEX-6P-2 (GE Healthcare) expression vectors for the expression of GST-CagA-2CM<sup>W</sup> from *H. pylori* 26695 strain and GST-CagA-4CM<sup>W</sup> from NCTC11637 strain have been described previously<sup>12,13</sup>. CM deletions and mutations were introduced by site-directed mutagenesis. In addition to a GST tag on their N-termini, all bacterial constructs used in this study contain a 6xHis tag on the C-termini for purification purposes. GST-PAR1b (39–364) and GST-PAR1b (39–364)-FLAG expression vectors were made by fusing a 6xHis tag sequence or a FLAG-6xHis tag sequence, respectively, to the 3' end of the *PAR1b* (39–364) fragment by PCR and then inserting these fragments into the pGEX-6P-3 (GE Healthcare) expression vector. Human *PAR1b* (GenBank: NM\_017490.3) was used as template for PCR.

**Antibodies.** Anti-FLAG monoclonal antibody (M2, Sigma), anti-HA monoclonal antibodies (3F10, Roche; 6E2, Cell Signaling Technology) and anti-actin polyclonal antibody (C-11, Santa Cruz Biotechnology) were used as primary antibodies for immunoblotting. Anti-HA (C29F4, Cell Signaling Technology) was used as primary antibody for immunostaining. Horseradish peroxidase-linked anti-mouse (GE Healthcare), anti-rat (GE Healthcare) and anti-goat (Santa Cruz Biotechnology) antibodies were used as secondary antibodies in immunoblotting. Alexa Fluor 546 anti-rabbit antibody (Invitrogen) was used in immunostaining as secondary antibody.

**Protein expression and purification.** Protein expression and purification of CagA and the GST control proteins were performed as previously described<sup>13</sup>. For the expression and purification of PAR1b (39–364) and PAR1b (39–364)-FLAG, *E. coli* BL21 harboring these expression vectors were grown at 37 °C in LB broth supplemented with ampicillin until OD<sub>600</sub> reached 0.4–0.6. Protein expression was induced with 25 μM IPTG for 24 h at 18 °C, after which cells were washed and pelleted. Cells were lysed by sonication in 50 mM HEPES, 300 mM NaCl, 5% glycerol, 10 mM imidazole (adjusted to pH7.2 with NaOH) and centrifuged for 15 min at 15,000 × g to clear the lysate of cell debris. The cleared lysate was then mixed with Ni-NTA Agarose beads (Qiagen), and the beads were washed with 50 mM HEPES, 300 mM NaCl, 5% glycerol, 20 mM imidazole (adjusted to pH7.2 with NaOH). The protein was eluted with 50 mM HEPES, 300 mM NaCl, 5% glycerol, 250 mM imidazole (adjusted to pH7.2 with HCl). 2 mM DTT was added to the eluate, followed by mixing with Glutathione Sepharose 4B beads (GE Healthcare). Beads were washed with GST-binding buffer (50 mM Tris-Cl, pH7.3, 150 mM NaCl, 5 mM EDTA, 2 mM DTT) and the GST tag was subsequently cleaved-off by PreScission Protease (GE Healthcare) in 50 mM Tris-Cl, pH7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT overnight. All PAR1b (39–364) purification steps were performed either on ice or at 4 °C.

**GST pull-down assay.** GST-fused proteins were bound to Glutathione Sepharose 4B beads as described above, and mixed for 1 h at 4 °C with purified proteins in HBS-P buffer (10 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 0.05% Tween-20) or GST-binding buffer + 0.01% Triton X-100. Beads were then washed five times with the binding buffer, resolved on SDS-PAGE gel and stained with Coomassie Brilliant Blue (CBB) for visualization.

**Size-exclusion chromatography (SEC).** Protein samples were subjected to analysis by Superdex 200 10/300 GL filtration column (GE Healthcare) on an AKTA avant 25 system (GE Healthcare) at 4 °C using HBS-P as running buffer, at a flow rate of 0.5 ml/min. Fractionated samples were resolved by SDS-PAGE, stained with CBB and quantified using the LAS-4000 Luminescent Image Analyzer System (Fujifilm).

**Analytical ultracentrifugation (AUC).** Protein samples were dialysed against 10 mM HEPES pH 7.7, 150 mM NaCl, 1 mM EDTA and the concentration of the sample was adjusted by using the dialysate, which was also used as a reference solution. Sedimentation velocity experiments were carried out at 20 °C using a ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter). Concentration gradients were measured with UV absorption at 280 nm with no time interval. The partial specific volumes of the proteins, the buffer density and viscosity were calculated by Sednterp<sup>56</sup>. The distribution functions of sedimentation coefficients,  $c(s)$ , were calculated using the SEDFIT program<sup>57</sup>, assuming that the frictional ratio was common to all the molecular species.  $c(s)$  was converted to  $c(M)$ , the distribution of the molecular weights, based on the Svedberg equation, which was implemented in SEDFIT.

**Quantitative GST pull-down assay.** Quantitative GST pull-down assay was designed as previously described<sup>44</sup>. In brief, to create a range of GST-CagA concentrations, different amounts of GST-CagA or GST bound Glutathione Sepharose 4B beads were aliquoted into reaction tubes. Unbound beads were then added to each tube accordingly, so that the total volume of beads per tube would be equal. Beads were mixed with PAR1b (39–364)-FLAG in GST-binding buffer containing 0.01% Triton-X for 20 min at 4 °C and pelleted in a standard microcentrifuge at 12,000 rpm for 2 min at 4 °C. Instead of quantifying the pellet bound fraction, which would require washing of the beads that would disturb the equilibrium of the interacting proteins, the unbound PAR1b (39–364)-FLAG was quantified by analysing the supernatant of the reaction by SDS-PAGE, followed by CBB staining or immunoblotting and quantification on a LAS-4000 Luminescent Image Analyzer System (Fujifilm). Bound PAR1b (39–364)-FLAG was determined by subtracting the unbound PAR1b (39–364)-FLAG from the total. Non-specific binding to GST-control was subtracted from bound PAR1b (39–364)-FLAG to determine the specific binding. The concentrations of GST-CagA and GST were determined by running a small sample of beads on an SDS-PAGE gel against known protein standards and quantifying the CBB-stained bands.  $K_D$  values were calculated by non-regression analysis using the GraphPad Prism 6 software (GraphPad Software). Non-linear regression analysis of the relationship between CM number and  $K_A$  values were also performed using the GraphPad Prism 6 software.

**Mammalian expression vectors.** The pCDH-EF1 vector was made by excising the CMV promoter and the puromycin resistance gene from the CD510B-1 lentivector (System Biosciences) and introducing an MCS downstream of the EF-1 promoter so that the inserted gene would be expressed under the control of the EF-1 promoter. pCDH-EF1-CagA- $\Delta$ CM, -2CM<sup>W</sup> and -4CM<sup>W</sup> were generated by inserting chemically synthesized humanized *cagA* constructs into pCDH-EF1. Humanized *cagA*-1CM<sup>W2</sup> was generated by deleting the first CM motif from humanized *cagA*-2CM<sup>W</sup> by site-directed mutagenesis. *cagA*-1CM<sup>E</sup> was also generated by site-directed mutagenesis using *cagA*-1CM<sup>W2</sup> as template. The mutagenized genes were inserted into pCDH-EF1 to make pCDH-EF1-CagA-1CM<sup>W2</sup> and pCDH-EF1-CagA-1CM<sup>E</sup>. These pCDH-EF1-CagA vectors were used to generate pseudoviral particles based on the Lentiviral Expression System (System Biosciences).

**AGS culture and transfection.** AGS human gastric epithelial cells were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum at 37 °C in 5% CO<sub>2</sub>. For microscopy, AGS cells were seeded in 8-well chamber wells at  $0.2 \times 10^5$  cells/well and after 12 h, transfected with 0.24  $\mu$ g of pCDH plasmids using Lipofectamine 2000 (Invitrogen). Cells were fixed for stress fiber analysis 24 h after transfection.

**Immunoblotting.** Preparation of AGS total cell lysates and immunoblotting were performed as described previously<sup>46</sup>.

**Immunostaining and microscopy.** AGS cells were fixed and permeabilized as described previously<sup>12</sup>. Cells were treated with primary antibodies followed by visualization with Alexa Fluor-conjugated secondary antibodies (Invitrogen). F-actin was stained with Alexa Fluor 488-conjugated phalloidin (Invitrogen) and nuclei were stained with DAPI (Wako). The focal plane of the confocal microscope was set at the cell bottom. Images were acquired using an Olympus FV1200 laser confocal microscope. Images were analysed using the Image J software (NIH)<sup>58</sup>.

**Quantification and statistics for stress fiber assay.** AGS cells, which displayed at least one stress fiber that transected the middle 50% of the cell body or the nucleus, were considered stress fiber-positive. The percentage of cells with stress fiber (mean  $\pm$  range %) was calculated from three independent experiments. The repeated measures one-way ANOVA and post-hoc Tukey's test were performed using the GraphPad Prism 6 software (GraphPad Software) based on an assumption that the data are normally distributed. Alpha level was set at 0.05 (two-tailed).



**Transepithelial electrical resistance (TER).** MDCK cells were allowed to form polarized monolayers for three days on transwell filters (#3460, Corning) and then infected with pseudoviral particles from the basal side at 200 MOI (multiplicity of infection) for 24 hours in the presence of 8 µg/ml polybrene. Resistance of samples were measured using Millicell ERS-2 (Millipore) and blank values were subtracted off to calculate the resistance of cell monolayers.

## References

1. Peek, R. M. & Blaser, M. J. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat. Rev. Cancer* **2**, 28–37 (2002).
2. Hatakeyama, M. *Helicobacter pylori* CagA and gastric cancer: A paradigm for hit-and-run carcinogenesis. *Cell Host Microbe* **15**, 306–316 (2014).
3. Nomura, A. *et al.* *Helicobacter pylori* infection and gastric carcinoma among Japanese Americans in Hawaii. *N. Engl. J. Med.* **325**, 1132–1136 (1991).
4. Parsonnet, J. *et al.* *Helicobacter pylori* infection and the risk of gastric carcinoma. *N. Engl. J. Med.* **325**, 1127–1131 (1991).
5. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
6. Bridge, D. R. & Merrell, D. S. Polymorphism in the *Helicobacter pylori* CagA and VacA toxins and disease. *Gut Microbes* **4**, 101–117 (2013).
7. Shiota, S., Suzuki, R. & Yamaoka, Y. The significance of virulence factors in *Helicobacter pylori*: Virulence factors in *H. pylori*. *J. Dig. Dis.* **14**, 341–349 (2013).
8. Hatakeyama, M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nat. Rev. Cancer* **4**, 688–694 (2004).
9. Hatakeyama, M. Linking epithelial polarity and carcinogenesis by multitasking *Helicobacter pylori* virulence factor CagA. *Oncogene* **27**, 7047–7054 (2008).
10. Odenbreit, S. *et al.* Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* **287**, 1497–1500 (2000).
11. Stein, M., Rappuoli, R. & Covacci, A. Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after cag-driven host cell translocation. *Proc. Natl. Acad. Sci.* **97**, 1263–1268 (2000).
12. Murata-Kamiya, N., Kikuchi, K., Hayashi, T., Higashi, H. & Hatakeyama, M. *Helicobacter pylori* exploits host membrane phosphatidylserine for delivery, localization, and pathophysiological action of the CagA oncoprotein. *Cell Host Microbe* **7**, 399–411 (2010).
13. Hayashi, T. *et al.* Tertiary structure-function analysis reveals the pathogenic signaling potentiation mechanism of *Helicobacter pylori* oncogenic effector CagA. *Cell Host Microbe* **12**, 20–33 (2012).
14. Selbach, M., Moese, S., Hauck, C. R., Meyer, T. F. & Backert, S. Src is the kinase of the *Helicobacter pylori* CagA protein *in vitro* and *in vivo*. *J. Biol. Chem.* **277**, 6775–6778 (2002).
15. Stein, M. *et al.* A c-Src/Lyn kinases activate *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol. Microbiol.* **43**, 971–980 (2002).
16. Higashi, H. *et al.* SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* **295**, 683–686 (2002).
17. Tsutsumi, R., Higashi, H., Higuchi, M., Okada, M. & Hatakeyama, M. Attenuation of *Helicobacter pylori* CagA-SHP-2 signaling by interaction between CagA and C-terminal Src kinase. *J. Biol. Chem.* **278**, 3664–3670 (2003).
18. Suzuki, M. *et al.* 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 (2005).
19. Higashi, H. *et al.* Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc. Natl. Acad. Sci.* **99**, 14428–14433 (2002).
20. Yamaoka, Y. *et al.* Relationship between the cagA 3' repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. *Gastroenterology* **117**, 342–349 (1999).
21. Azuma, T. *et al.* Association between diversity in the Src homology 2 domain-containing tyrosine phosphatase binding site of *Helicobacter pylori* CagA protein and gastric atrophy and cancer. *J. Infect. Dis.* **189**, 820–827 (2004).
22. Ren, S., Higashi, H., Lu, H., Azuma, T. & Hatakeyama, M. Structural basis and functional consequence of *Helicobacter pylori* CagA multimerization in cells. *J. Biol. Chem.* **281**, 32344–32352 (2006).
23. Saadat, I. *et al.* *Helicobacter pylori* CagA targets PARI/MARK kinase to disrupt epithelial cell polarity. *Nature* **447**, 330–333 (2007).
24. Nesić, D. *et al.* *Helicobacter pylori* CagA inhibits PARI-MARK family kinases by mimicking host substrates. *Nat. Struct. Mol. Biol.* **17**, 130–132 (2010).
25. Kempthues, K. J., Priess, J. R., Morton, D. G. & Cheng, N. Identification of genes required for cytoplasmic localization in early *C. elegans* embryos. *Cell* **52**, 311–320 (1988).
26. Drewes, G. *et al.* Microtubule-associated protein/microtubule affinity-regulating kinase (p110mark): A novel protein kinase that regulates tau-microtubule interactions and dynamic instability by phosphorylation at the Alzheimer-specific site serine 262. *J. Biol. Chem.* **270**, 7679–7688 (1995).
27. Drewes, G., Ebnet, A., Preuss, U., Mandelkow, E.-M. & Mandelkow, E. MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption. *Cell* **89**, 297–308 (1997).
28. Böhm, H., Brinkmann, V., Drab, M., Henske, A. & Kurzchalia, T. V. Mammalian homologues of *C. elegans* PAR-1 are asymmetrically localized in epithelial cells and may influence their polarity. *Curr. Biol.* **7**, 603–606 (1997).
29. Hurov, J. B., Watkins, J. L. & Piwnicka-Worms, H. Atypical PKC phosphorylates PAR-1 kinases to regulate localization and activity. *Curr. Biol.* **14**, 736–741 (2004).
30. Suzuki, A. *et al.* aPKC acts upstream of PAR-1b in both the establishment and maintenance of mammalian epithelial polarity. *Curr. Biol.* **14**, 1425–1435 (2004).
31. Lu, H.-S. *et al.* Structural and functional diversity in the PAR1b/MARK2-binding region of *Helicobacter pylori* CagA. *Cancer Sci.* **99**, 2004–2011 (2008).
32. Lu, H., Murata-Kamiya, N., Saito, Y. & Hatakeyama, M. Role of partitioning-defective 1/microtubule affinity-regulating kinases in the morphogenetic activity of *Helicobacter pylori* CagA. *J. Biol. Chem.* **284**, 23024–23036 (2009).
33. Cohen, D., Fernandez, D., Lázaro-Diéguéz, F. & Müsch, A. The serine/threonine kinase Par1b regulates epithelial lumen polarity via IRSp53-mediated cell–ECM signaling. *J. Cell Biol.* **192**, 525–540 (2011).
34. Yamahashi, Y., Saito, Y., Murata-Kamiya, N. & Hatakeyama, M. Polarity-regulating kinase partitioning-defective 1b (PAR1b) phosphorylates guanine nucleotide exchange factor H1 (GEF-H1) to regulate RhoA-dependent actin cytoskeletal reorganization. *J. Biol. Chem.* **286**, 44576–44584 (2011).
35. Tojkander, S., Gateva, G. & Lappalainen, P. Actin stress fibers – assembly, dynamics and biological roles. *J. Cell Sci.* **125**, 1855–1864 (2012).
36. Vallenius, T. Actin stress fiber subtypes in mesenchymal-migrating cells. *Open Biol.* **3**, 130001 (2013).
37. Kersulyte, D. *et al.* *Helicobacter pylori* from Peruvian Amerindians: Traces of human migrations in strains from remote Amazon, and genome sequence of an Amerind strain. *PLoS ONE* **5**, e15076 (2010).

38. Mane, S. P. *et al.* Host-interactive genes in Amerindian *Helicobacter pylori* diverge from their Old World homologs and mediate inflammatory responses. *J. Bacteriol.* **192**, 3078–3092 (2010).
39. Suzuki, M. *et al.* Attenuated CagA oncoprotein in *Helicobacter pylori* from Amerindians in Peruvian Amazon. *J. Biol. Chem.* **286**, 29964–29972 (2011).
40. Hashi, K. *et al.* Natural variant of the *Helicobacter pylori* CagA oncoprotein that lost the ability to interact with PAR1. *Cancer Sci.* **105**, 245–251 (2014).
41. Zeaiter, Z. *et al.* 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 (2008).
42. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
43. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server In *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005).
44. Pollard, T. D. A guide to simple and informative binding assays. *Mol. Biol. Cell* **21**, 4061–4067 (2010).
45. Higashi, H. *et al.* EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J. Biol. Chem.* **280**, 23130–23137 (2005).
46. Nagase, L., Hayashi, T., Senda, T. & Hatakeyama, M. Dramatic increase in SHP2 binding activity of *Helicobacter pylori* Western CagA by EPIYA-C duplication: its implications in gastric carcinogenesis. *Sci. Rep.* **5**, 15749 (2015).
47. Tan, S., Tompkins, L. S. & Amieva, M. R. *Helicobacter pylori* usurps cell polarity to turn the cell surface into a replicative niche. *PLoS Pathog.* **5**, e1000407 (2009).
48. Tan, S., Noto, J. M., Romero-Gallo, J., Peek, R. M. & Amieva, M. R. *Helicobacter pylori* perturbs iron trafficking in the epithelium to grow on the cell surface. *PLoS Pathog.* **7**, e1002050 (2011).
49. Noto, J. M. *et al.* Iron deficiency accelerates *Helicobacter pylori*-induced carcinogenesis in rodents and humans. *J. Clin. Invest.* **123**, 479–492 (2013).
50. Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D. & Nussinov, R. Extended disordered proteins: targeting function with less scaffold. *Trends Biochem. Sci.* **28**, 81–85 (2003).
51. Cortese, M. S., Uversky, V. N. & Dunker, A. K. Intrinsic disorder in scaffold proteins: Getting more from less. *Prog. Biophys. Mol. Biol.* **98**, 85–106 (2008).
52. Domínguez-Bello, M. G. *et al.* Amerindian *Helicobacter pylori* strains go extinct, as European strains expand their host range. *PLoS ONE* **3**, e3307 (2008).
53. Nagase, L., Murata-Kamiya, N. & Hatakeyama, M. Potentiation of *Helicobacter pylori* CagA protein virulence through homodimerization. *J. Biol. Chem.* **286**, 33622–33631 (2011).
54. Croft, D. R. *et al.* Conditional ROCK activation *in vivo* induces tumor cell dissemination and angiogenesis. *Cancer Res.* **64**, 8994–9001 (2004).
55. Jou, T.-S. & Nelson, W. J. Effects of regulated expression of mutant RhoA and Rac1 small GTPases on the development of epithelial (MDCK) cell polarity. *J. Cell Biol.* **142**, 85–100 (1998).
56. Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (eds Harding, S. E., Rowe, A. J. & Horton, J. C.) 90–125 (R Soc Chem, 1992).
57. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
58. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).

## Acknowledgements

We thank members of the Hatakeyama Lab, especially Naoko Kamiya, for technical guidance and useful discussions. This work was supported by CREST from Japan Science and Technology (Grant Number 12020000396) and by Grant-in-Aid for Scientific Research on Innovative Area from the Ministry of Education, Culture, Sports, and Technology of Japan (MEXT KAKENHI Grant Number 25250016).

## Author Contributions

Designed research: H.N. and M.H. Performed research: H.N., T.H. and F.A. Analyzed data: H.N., T.H., F.A. and T.S. Wrote the paper: H.N. and M.H.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Nishikawa, H. *et al.* Impact of structural polymorphism for the *Helicobacter pylori* CagA oncoprotein on binding to polarity-regulating kinase PAR1b. *Sci. Rep.* **6**, 30031; doi: 10.1038/srep30031 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>