Helicobacter pylori CagA Inhibits PAR1/MARK Family Kinases by Mimicking Host Substrates

Dragana Nešić[1], Marshall C. Miller[1], Zachary T. Quinkert[2], Markus Stein[3], Brian T. Chait[2], and C. Erec Stebbins[1]†

Supplementary Discussion

Establishing the Stoichiometry of the CagA-MARK2 Interaction

Because the MKI sequence occurs twice in the crystallized construct, and only the amino acids common to both repeats are visible, it is unclear from the crystal structure which of the two possible peptide regions is binding (the first or second repeat of the MKI sequence). This analysis was further complicated by the presence of four independent molecules of MARK2 in the asymmetric unit of the crystals, each with the peptide bound. Mutating or deleting either one of the two repeats alone (leaving the other intact) resulted in CagA remaining bound to MARK2 (Fig. 3a and Supplementary Fig. 2), suggesting that both MKI repeats are able to interact with the kinase. This was perhaps unsurprising as East Asian CagA proteins contain only one MKI sequence[1]. This also implied that a single CagA polypeptide with multiple MKI sequences could potentially bind to multiple MARK2 molecules, raising the question of stoichiometry of the complex.

We addressed this issue through a combination of gel filtration chromatography and native mass spectrometry. These results clearly demonstrate that each MKI sequence is bound by a molecule of MARK2. Gel filtration profiles indicated that binding of MARK2 to CagA

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

induces dimerization of MARK2 (Supplementary Fig. 2), but due to limitations of gel filtration chromatography and the size of the purified construct, it was impossible to distinguish between complexes of the form $(CagA)_2(MARK2)_2$ (1:1 molar ratio) and $(CagA)_1(MARK2)_2$ (1:2 molar ratio).

To resolve this ambiguity, we analyzed the CagA-MARK2 complex by native electrospray ionization mass spectrometry[2,3]. The high mass accuracy and resolution of native mass spectrometry allows for direct stoichiometry determination of the intact complex. The native mass spectrum of the wild type CagA-MARK2 complex (Supplementary Fig. 3) is dominated by the $(CagA)_1(MARK2)_2$ stoichiometric form. Although a small amount of $(CagA)_1(MARK2)_1$ as well as monomeric CagA and MARK2 is also observed, these likely arise from partial dissociation of the native complex in the ammonium acetate buffer solution required for mass spectrometry or during the electrospray ionization process – a conclusion supported by the absence of monomeric CagA and MARK2 in the size exclusion chromatogram. Additional support for our conclusion that the relevant stoichiometric form is $(CagA)_1(MARK2)_2$ is provided by the native mass spectrum of the complex of a mutant form of CagA and MARK2 (Supplementary Figs. 3 and 4), where the mutant has only a single functional MARK2 binding site (CagA(LRKL) with mutations L950G, R952G, K955G, and L955G). In this case, the native mass spectrum is dominated by $(CagA)_1(MARK2)_1$ and the $(CagA)_1(MARK2)_2$ stoichiometric form is entirely absent. Size exclusion chromatography of the CagA(LRKL)-MARK2 complex also indicates a shift towards lower molecular weight (Fig. 3a and Supplementary Fig. 2). Finally, we observed that the wild-type CagA-MARK2 stoichiometry of 1:2 was maintained over a concentration range of 10-100μM (Supplementary Fig. 5). These results strongly suggest that, at least in solution, each repeat of the MKI sequence is bound by a molecule of MARK2.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

Additional Remarks on Mutagenesis

Mutations of Phe948 to glycine moderately decreased binding of CagA to MARK2, but mutations of H953G and K955G had no effect in the pull-down assay. Supporting this data, East-Asian CagA subtypes contain glycine in the position that corresponds to Lys955 in Western CagA, and it has been reported that MARK2 binds more strongly to the East Asian CagA repeats region[4]. LD mutations (L248G, D251G) in MARK2 did not alter binding to CagA (in pulldown experiments, Fig. 3a, or ion-exchange or gel filtration chromatography), which may be explained by the fact that these interactions are dominated by main-chain interactions to CagA side-chains.

The F948G mutant, which has decreased binding to MARK2, was also a weaker inhibitor of MARK2 kinase activity (Fig. 3, b-d). This was even more prominent when non-activated MARK2 was used, where the F948G mutant appeared 10-20 times weaker an inhibitor than wild type or the K955G peptide (Fig. 3d).

Our mutagenesis results are consistent with the idea that one would expect a pseudosubstrate inhibitor to bind the active conformation of the kinase, and also expect a higher affinity inhibition with the phosphorylated (active) MARK2 than with the unphosphorylated kinase. It is likely that the high affinity of the interaction with the activated form of MARK2 explains the induction of an active-like state of the kinase in our crystals.

Additional Concluding Remarks

CagA is a central virulence factor of *H. pylori*, and is strongly tied epidemiologically to virulence and carcinogenesis. The only known substrate of the T4SS of the pathogen, this large protein has proven since its cloning in 1993[5,6] to be recalcitrant to structural studies. Here we

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

present the first structural data and mechanistic insight into the function of CagA, co-crystallizing the repeats domain with one of its many host cell targets, the MARK2 protein kinase.

The MKI motif occurs twice in the crystallized construct, and is contained in a sequence (termed the CM sequence for CagA Multimerization) that was recently shown to be important for CagA activity towards MARK2 [1,4,7]. Prior work using deletional analyses had suggested that the interaction of CagA with MARK2 involves a region of the kinase domain at the far C-terminal lobe (residues 250-276) [1]. However, the removal of 27 amino acids from the kinase fold, including a helix packing into the hydrophobic core, likely led to a destabilization of the structure, including the neighbouring region encompassing the substrate-binding site (Supplementary Fig. 6). In contrast, our studies clearly demarcate a different and more biologically relevant interaction surface (Supplementary Fig. 6).

It was also recently reported[8] that MKI (termed CRPIA in the report for Conserved Repeat responsible for Phosphorylation–Independent Activity) mediates sustained activation of Met/PI3K/Akt signalling that leads to the up-regulation of β-catenin and NFκB transcriptional activities. It is suggested that those signalling events contribute to the epithelial proliferative and proimmflamatory responses associated with Helicobacter-induced gastric diseases. It is not clear at this point if MKI specific inhibition of MARK2 could be linked to those events or if this is completely independent and new role of MKI sequence. Currently, therefore, there are three acronyms that essentially correspond to the same sequence in CagA (CM, CRPIA, and MKI). Because our structural data unambiguously demonstrates that the peptide is a PKI-like inhibitor, we prefer the designation “MKI,” which more specifically identifies the activity of this region.

Finally, MARK2 presents an attractive target for therapeutic intervention. Its possible contribution to neurodegenerative diseases through its phosphorylation of the MAP tau protein[9], as well as its apparent role in the regulation of metabolism, insulin secretion and body mass[10-12], imply that pharmacological inhibitors of MARK2 could serve as treatments for amyloid-like diseases as well as obesity and diabetes. The structure of CagA-MARK2 can, in this context, provide a blueprint for the rational design of therapeutic peptides and small molecules. This would provide a remarkable example of a bacterial virulence factor serving as a template for the development of novel chemotherapeutic interventions for human diseases.

Accession codes. Protein Data Bank: Coordinates for the CagA-MARK2 complex have been deposited with accession code 3IEC.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

Supplementary Methods

Purification of the *Helicobacter pylori* 26695 CagA and MARK2 complex. Full length *Helicobacter pylori 26695* CagA was cloned by PCR from genomic DNA (obtained from the American Type Culture Collection, reference ATCC#700392D) in pcDNA4/HisMax© TOPO® (Invitrogen) vector. All domain specific constructs (885-1006 and 885-981) used in this study are derived from that original plasmid by PCR and subcloned as an N-terminal hexahistidine fusion protein into an engineered version of the pCDFDuet-1 vector (Novagen), which contains two tandem hexahistidine tags separated by linker and followed by rhinovirus 3C protease recognition sequence. MARK2 (39-364) was subcloned by PCR from full length human MARK2 cDNA clone (obtained from ImaGenes, ref# IRAUp969E03113D). Un-tagged MARK2 was ligated into the second cloning site of modified pCDFDuet-1, downstream of the hexahistidine-tagged CagA, which has a separate T7 promoter and ribosomal biding site. Proteins were co-expressed in BL-21 *E. coli* following induction with 0.1mM IPTG at 26°C overnight. Un-tagged MARK2 and hexahistidine tagged CagA were co-purified on Ni-NTA Sepharose (Qiagen) in elution buffer consisting of 20mM HEPES pH 7.5, 200mM NaCl, and 250mM Imidazole. CagA was separated from the affinity tag by site-specific proteolytic cleavage with heaxhistidine tagged 3C protease (1:100), and the protease removed, as well as uncleaved CagA, by additional pass through Ni-NTA column. The CagA-MARK2 complex was further purified by anion-exchange (MonoQ 10/100 GL, 8 ml bed volume, GE Health) and cation-exchange (Fast Flow SP Sepharose, 20 ml bead volume, GE Health), and gel-filtration chromatography (Superdex 200, GE Healthcare) using an ÄKTA® FPLC (GE Health). Prior loading the ion- exchange columns, the buffer containing the protein complex was diluted to lower the salt concentration to 40mM NaCl. After equilibrating the columns with washing buffer

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

(20mM HEPES pH 7.5, 2.5mM dithiothreitol (DTT)), protein complexes were loaded and unbound material removed with 3 column volumes of washing buffer. Proteins that were bound to the columns were subjected to a gradient increase of salt concentration (from 0 to 500mM NaCl in 20mM HEPES pH 7.5, 2.5mM DTT using at least 20 column volumes). Even though CagA-MARK2 complexes bind very weakly to anion-exchange column, this was a critical step in obtaining the material for crystallization. CagA-MARK2 complexes bind much stronger to the cation-exchange column and elute at 200-230 mM NaCl. As a final step in the purification, and at the same time test of complex integrity, fractions eluted from SP Sepharose column that contained intact CagA-MARK2 complexes were pooled together, concentrated using Centricon® Plus-80 (Millipore), and run on gel filtration column (Superdex™ 200 HiLoad 16/60 (120ml) or Superdex™ 200 HR 10/30 (25ml) (GE Health). The gel filtration buffers contained 200mM NaCl, 20mM HEPES pH 7.5, 2.5mM dithiothreitol (DTT), 2ml sample was injected, and 2ml (from Superdex™ 200 HiLoad 16/60) or 1ml (from Superdex™ 200 HR 10/30) fractions were collected. The peak fractions were examined by SDS-PAGE, and proteins visualized by staining with Coomassie blue staining solution (50% methanol, 0.05% Coomassie brilliant blue R-50, 10% acetic acid in water) followed by destaining solution (5%methanol, 7% acetic acid in water).

For kinase studies, MARK2 (39-364) was expressed as N-terminal hexahistidine fusion protein and purified following the same protocol that was used for the purification of CagA-MARK2 complexes, except that anion-exchange step was skipped. During model building, it was discovered that there was a mutation in the sequence of MARK2 (R146W). While the structure was refined with this mutation, the wild-type protein was used in all remaining studies (binding assays, kinases assays, etc), although studies with this mutant showed that the amino

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

acid change had no discernable effect on the biochemical properties of the protein, either in purification, CagA binding, or biochemistry (data not shown).

Crystallization and structural determination. For crystallization, the purified complex of the CagA (885-1006) and MARK2 (39-364) was reductively methylated according to published protocols[13]. Methylated protein complexes were concentrated to 10-25mg/ml in a buffer containing 20mM HEPES pH 7.5, 200mM NaCl, and 2.5mM dithiothreitol (DTT) by gel filtration. Initial crystals were grown by vapor diffusion using hanging drops formed from mixing a 2:1 volume ratio of the protein complex with an equilibration buffer consisting of 22.5-25% polyethylene glycol (PEG) molecular weight 3350 Da, 300 mM Li_2SO_4, and 100 mM Bis-Tris pH 5.8-6.2, at 23ºC. Higher quality crystals were obtained by micro-seeding and addition of 5% PEG 400 as an additive. For cryoprotection crystals where transferred directly into a buffer with a 25% PEG 3350 Da, 5-10% PEG 400 Da, 300mM Li_2SO_4, 100mM Bis-Tris pH 5.8-6.2, 7.5% glycerol, and flash-cooled immediately afterward to -160°C.

Data were collected at Brookhaven National Synchrotron Light Source beamline X29. The structure was solved by molecular replacement using the program PHASER and the model of MARK2 (PDB ID 2HAK). Following several cycles of refinement with REFMAC5[14] and manual adjustment of the model with O[15], the electron density of the peptide was clearly visible and interpretable. The peptide was built manually with O[15], and the entire structure (four copies of MARK2 and the CagA 14-mer in the asymmetric unit) refined with REFMAC5. Final statistics of the refinement include an R/Rfree of 21%/25% with 90% of the residues in the most favored region of the Ramachandran plot.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

Mutagenesis. The deletion mutants were generated by PCR and cloned into the same vector as wild type constructs, modified pCDFDuet-1. The amino acid substitutions were introduced by PCR using *Pfu* DNA polymerase (Stratagene) and primers containing the appropriate base changes. The template plasmid was subsequently removed by digestion with DpnI restriction enzyme prior to transformation. All the mutations were verified by sequencing. Purification of the mutant proteins and protein complexes was performed as previously described for wild type proteins.

Binding Assay and Western blotting. Pull-down on Ni-NTA sepharose columns was used to determine binding of wild type or mutant hexahistidine-tagged Cag A (885-981) to wild type or mutant MARK2 (39-364). The soluble fraction of bacterial cell extracts co-expresssing His-CagA and MARK2 was loaded to the Ni-NTA columns in a binding buffer (20mM HEPES pH 7.5, 200mM NaCl), washed with the binding buffer plus gradual increase of imidazole (5mM, 10mM, 30mM and 50mM imidazole) and complexes eluted with elution buffer (binding buffer with 100-250mM imidazole). The initial assessment of complex formation was performed by gradient (4-12%) SDS-PAGE analysis of eluted material. If both proteins were present in the eluted fraction, it was further subjected to gradient ion-exchange (Fast Flow SP Sepharose, 20 ml bead volume, GE Health; 0 to 500mM NaCl in 20mM HEPES pH 7.5, 2.5mM DTT during 150 min at 3ml/min flow rate), and gel-filtration chromatography (Superdex 200, GE Healthcare; in 20mM HEPES pH 7.5, 200mM NaCl, 2.5mM DTT) in order to examine complex stability. Western blotting of total bacterial cell extract was performed to determine the level of expression of wild type and mutant hexahistidine-tagged CagA and MARK2. Bacterial cultures were lysed in sample buffer (50mM Tris-Cl pH6.8, 2% SDS 10% glycerol, 0.1% bromophenol blue,

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

100mM DTT) and proteins separated by SDS-PAGE. The proteins were then transferred to nitrocellulose membrane and His-CagA was blotted with anti-polyhistidine antibody directly conjugated with peroxidase (SIGMA, #A 7058; 1:5000 dilution in 5% milk TBST-20mM Tris pH 7.4, 400mM NaCl, 0.2% Tween-20). MARK-2 was detected with anti-phospho-MARK family antibody (Cell Signaling, #4836; 1:1000 dilution in 5% milk TBST), followed by goat anti-rabbit conjugated with peroxidase (Biorad #166-2408; 1:10000 in 5% milk TBST). Phosphorylated MARK2 was present within bacteria, which we explain as most likely due to autophosphorylation.

Kinase assays and IC_{50} determination. Kinase activity of MARK2 was measured using the luminescent kinase assay kit (Kinase-Glo®-Plus, Promega) according to manufacturer's recommendations. Kinase reactions were performed at 30°C for 30 minutes in a kinase reaction buffer (50mM Tris·Cl pH 7.5, 5mM $MgCl_2$, 2mM EGTA, 0.5mM PMSF, 0.5mM DTT, 0.5mM benzamidine) containing 5μg of MARK2, 10μM ATP, TR1 tau (KNVKSKIGSTENLK) peptide as a substrate (500μM) together with wild type (GFPLKRHDKVDDLSKVG) or mutant CagA synthetic MKI peptides at indicated concentrations. All synthetic peptides were of >95% purity (GeneScript Corporation). When MARK2 was activated by phosphorylation with MARKK (1:20 ratio of MARK2 to N-terminally GST tagged TAOK1 purchased from SignalChem), the kinase reaction was carried out under the same conditions (50mM Tris·Cl pH 7.5, 5mM $MgCl_2$, 2mM EGTA, 0.5mM PMSF, 0.5mM DTT, 0.5mM benzamidine, 10μM ATP) but without tau substrate and MKI peptides. Following 30 minutes incubation at 30°C, tau and MKI peptides were added and the reaction continued for an additional 30 minutes. Using this protocol, ATP consumption in the first reaction was equal in all reactions, and could be easily subtracted by using control reaction without a second substrate (tau). CagA MKI peptides had no effect on the

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

phosphorylation of MARK2 by MARKK (data not shown). After addition of Kinase-Glo® Plus Reagent luminescence was read using SpectraMax GeminiXS spectrofluorometer. In order to determine IC_{50} and K_i of MKI peptides the reaction was optimized to contain 2μg MARK2 (amount that provides luminescence values in the linear range of the kinase titration curve) and tau substrate concentration around Kd values (150μM - in the experiments with previously activated MARK2, or 200μM tau- in experiments without prior activation of MARK2). Kd values of tau peptide substrate were experimentally determined using the same luminescent kinase assay, and were 169.3 μM for MARKK activated MARK2 and 172 μM for MARK2 without prior activation with activating kinase. Curve fitting and IC_{50} and K_i calculations were performed using GraphPadPrism® 4 sigmoidal dose response software.

Mass Spectrometry

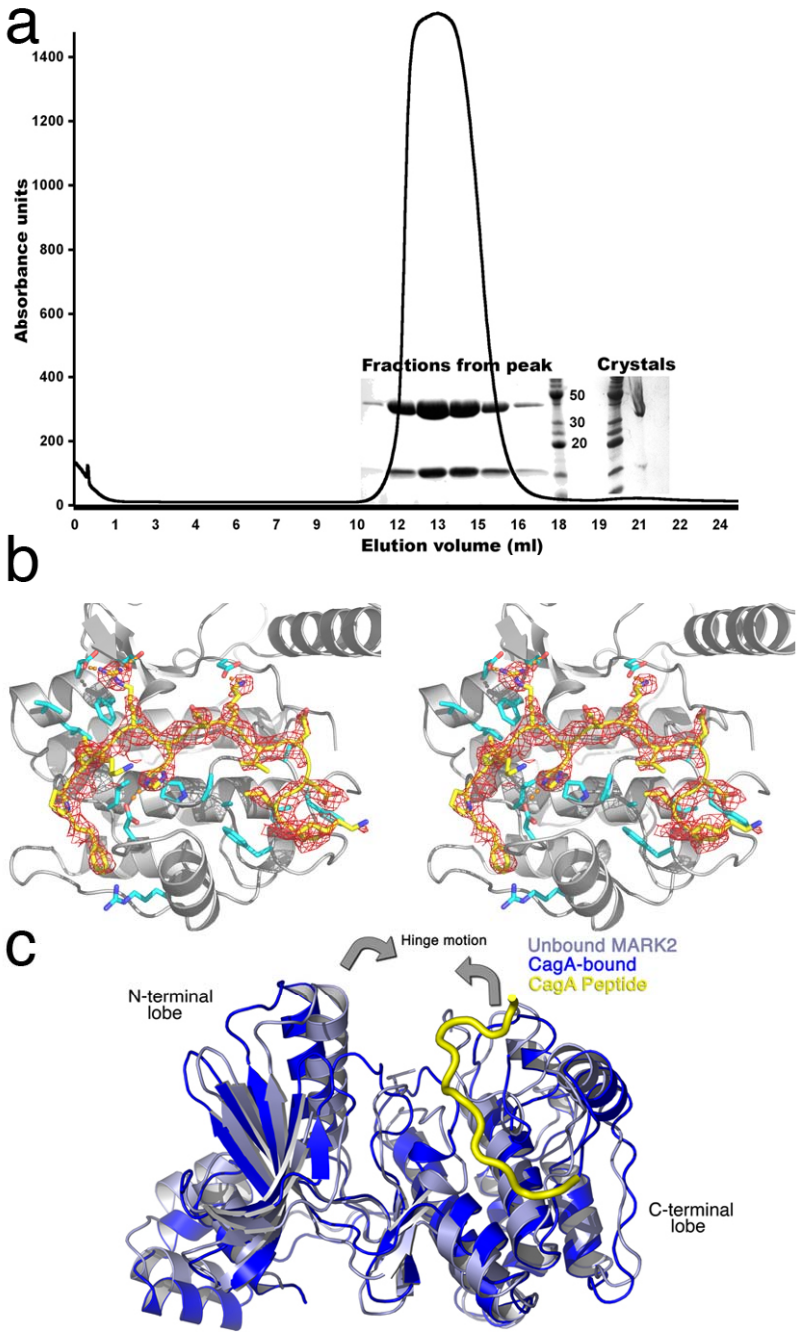
Samples were prepared for native MS by buffer exchanging 50μL of CagA + MARK2 (100μM stored in 20mM HEPES pH 7.5, 200mM NaCl, 2.5mM DTT) into 300mM ammonium acetate buffer (pH 7.4) via a gel filtration spin-column (Biorad, Micro Bio-spin 6, Cat. #732-6221). An aliquot of buffer-exchanged sample was diluted to the desired concentration by addition of 300mM ammonium acetate buffer, and ~2μL of the resulting solution was injected into a gold-sputtered borosilicate capillary (Waters, #M956232AD1-S) using GE tips (Eppendorf, Cat. #022351656). The capillary was spun briefly on a benchtop centrifuge to insure that the sample was located at the capillary tip. The capillary was mounted on the nano-electrospray source of a Synapt HDMS (Waters) and running conditions were optimized (see Supplementary Tables S2-S4). Approximately 300 scans were combined for each native mass spectrum using MassLynx v4.1 software (Waters). Several replicate measurements were made for each spectrum.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1705

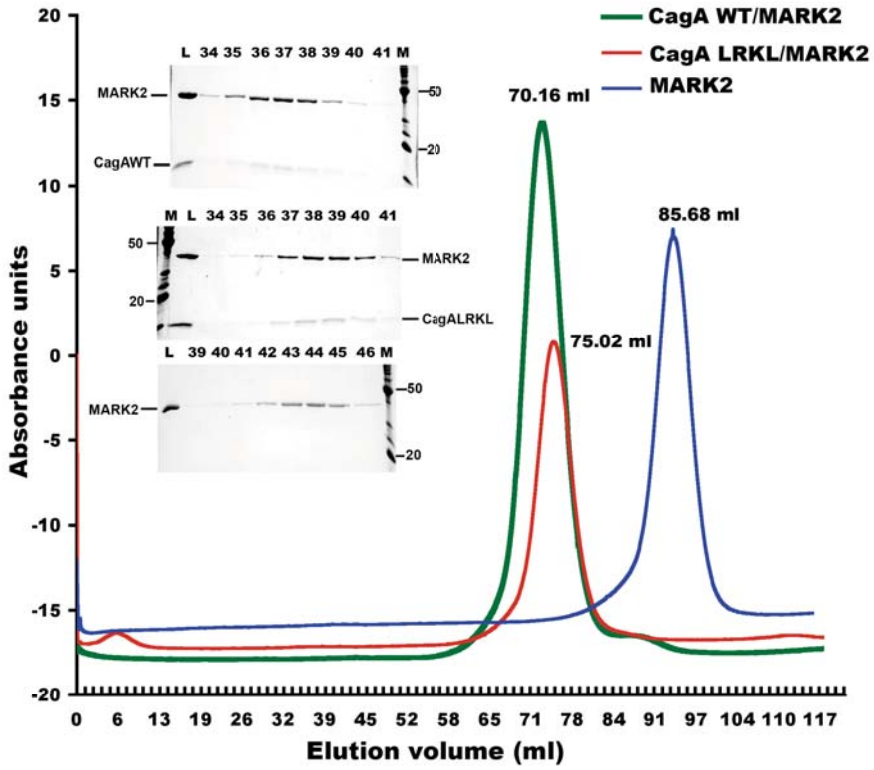
Supplementary References

1. Saadat, I. et al. Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature* **447**, 330-3 (2007).
2. Sharon, M. & Robinson, C.V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* **76**, 167-93 (2007).
3. Heck, A.J. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **5**, 927-33 (2008).
4. Lu, H.S. et al. Structural and functional diversity in the PAR1b/MARK2-binding region of Helicobacter pylori CagA. *Cancer Sci* **99**, 2004-11 (2008).
5. Tummuru, M.K., Cover, T.L. & Blaser, M.J. Cloning and expression of a high-molecular-mass major antigen of Helicobacter pylori: evidence of linkage to cytotoxin production. *Infect Immun* **61**, 1799-809 (1993).
6. Covacci, A. et al. Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci U S A* **90**, 5791-5 (1993).
7. 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**, 32344-52 (2006).
8. Suzuki, M. et al. Helicobacter pylori CagA phosphorylation-independent function in epithelial proliferation and inflammation. *Cell Host Microbe* **5**, 23-34 (2009).
9. Drewes, G., Ebneth, 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).
10. Hurov, J. & Piwnica-Worms, H. The Par-1/MARK family of protein kinases: from polarity to metabolism. *Cell Cycle* **6**, 1966-9 (2007).
11. Hurov, J.B. et al. Loss of the Par-1b/MARK2 polarity kinase leads to increased metabolic rate, decreased adiposity, and insulin hypersensitivity in vivo. *Proc Natl Acad Sci U S A* **104**, 5680-5 (2007).
12. Jansson, D. et al. Glucose controls CREB activity in islet cells via regulated phosphorylation of TORC2. *Proc Natl Acad Sci U S A* **105**, 10161-6 (2008).
13. Rypniewski, W.R., Holden, H.M. & Rayment, I. Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-A resolution. *Biochemistry* **32**, 9851-8 (1993).
14. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55 (1997).
15. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2)**, 110-9 (1991).
16. Panneerselvam, S., Marx, A., Mandelkow, E.M. & Mandelkow, E. Structure of the catalytic and ubiquitin-associated domains of the protein kinase MARK/Par-1. *Structure* **14**, 173-83 (2006).

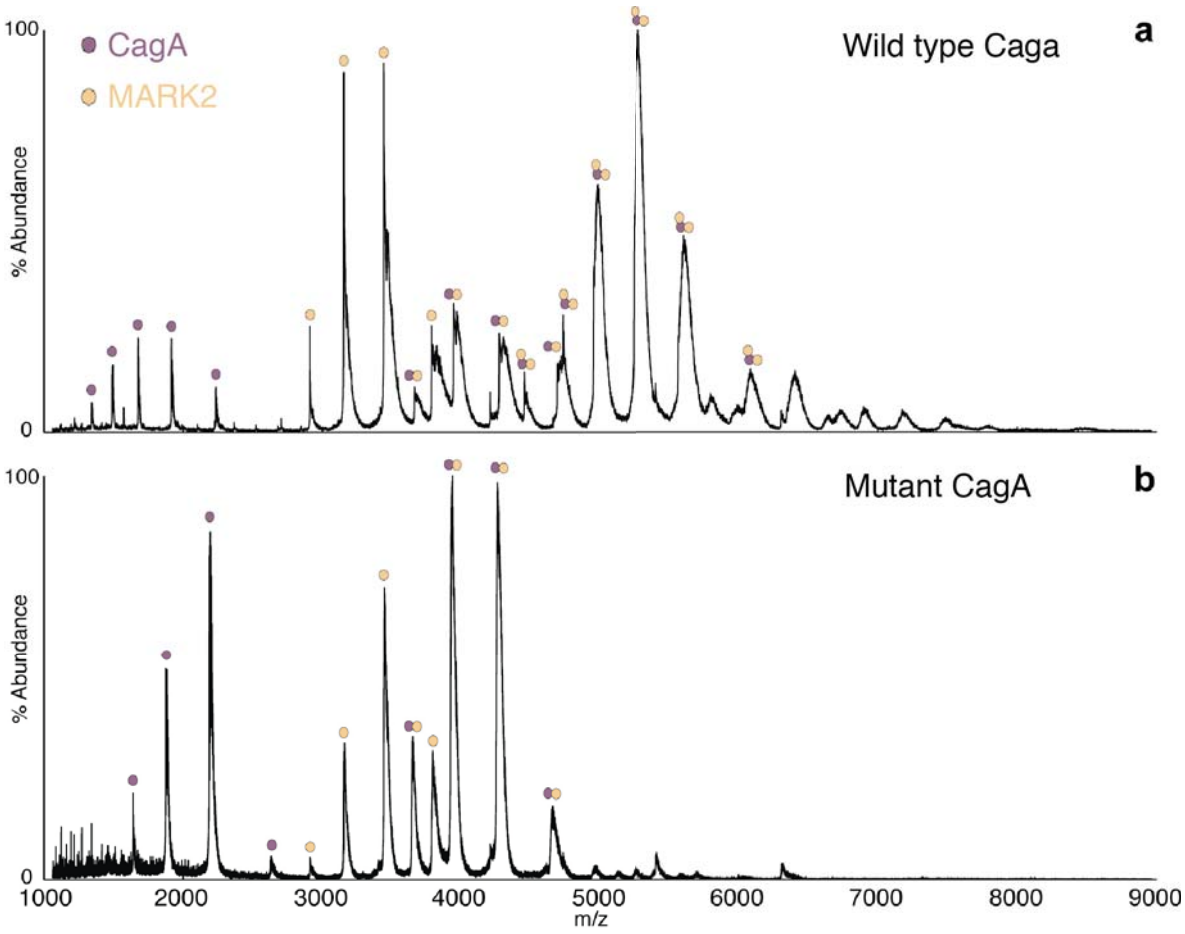
Supplemental Figures



Supplementary Fig. 1: Biochemistry and Structure (a) Size exclusion chromatography (gel filtration) of methylated Cag A (885-1005; 13kDa) and MARK2 (39-349; 37.5kDa) complex. Material eluted from this column was used to set up crystallization experiments. The SDS-PAGE on the right shows that both CagA (13kDa) and MARK2 (37.5kDa) are present inside crystallized material. The numbers by the molecular weight markers (extreme right and left of the two gels above) are in kDa. (b) Stereo diagram showing the CagA peptide (yellow) and MARK2 substrate binding site (grey/cyan). A 2fo-fc electron density map, calculated from model phases following the initial molecular replacement solution of MARK2 but before the CagA peptide was built, is shown at a sigma cutoff of 1.0 (red wire mesh). The CagA and MARK2 polypeptides are from the final, refined model. (c) Ribbon diagram showing a relative hinge-like motion between the unbound (blue-grey) and CagA bound (blue) crystal structures of MARK2. The CagA peptide is shown in yellow. The structure of MARK2 in complex with CagA is very similar to that of MARK2 alone[16], the two models aligning with a root-mean-square-deviation between Ca atoms of 2.3Å over 312 residues (aligning residues 51-363).

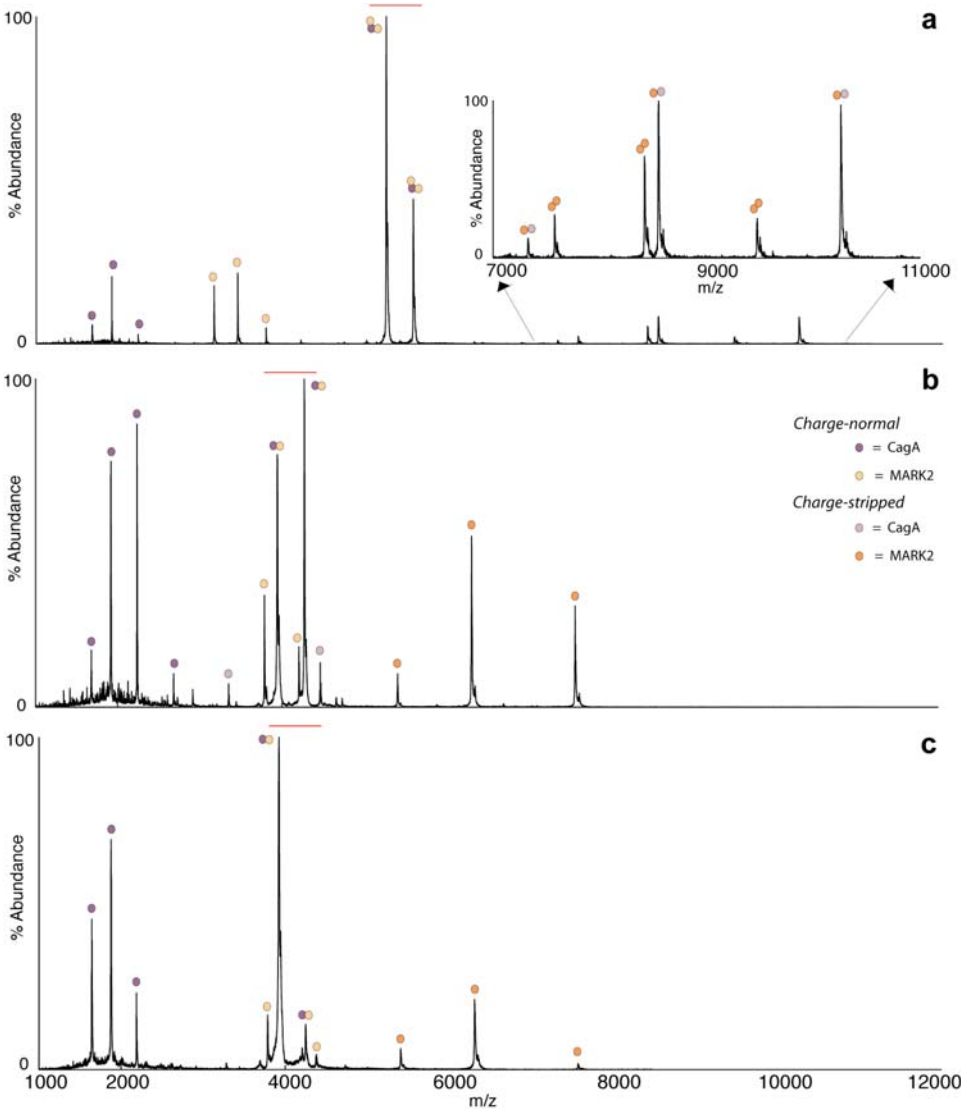


Supplementary Fig. 2: Comparison of gel filtration (Superdex 200) profiles of the complexes of MARK2 (39-349) with wild type CagA (885-1005), which has two MARK2 binding sites, and mutant LRKL Cag A (885-1005), which contains only one MARK2 binding site (LRKL- L950G, R952G, K955G, L959G). SDS-PAGE of eluted fractions is shown inside the graph. L-material loaded on gel filtration column; M-molecular weight marker in units of kDa.

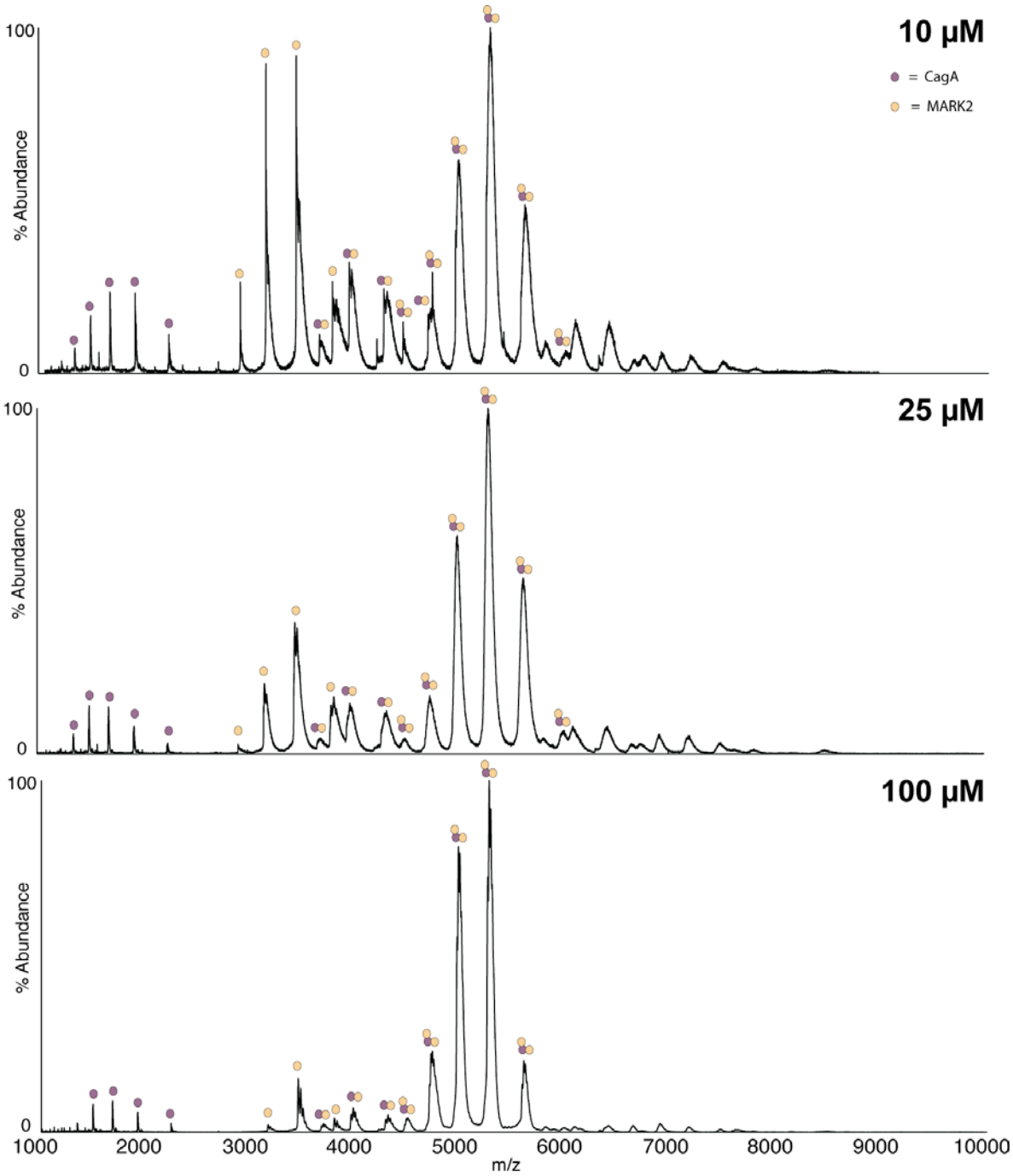


Supplementary Figure 3: CagA Binds MARK2 in a 1:2 stoichiometry

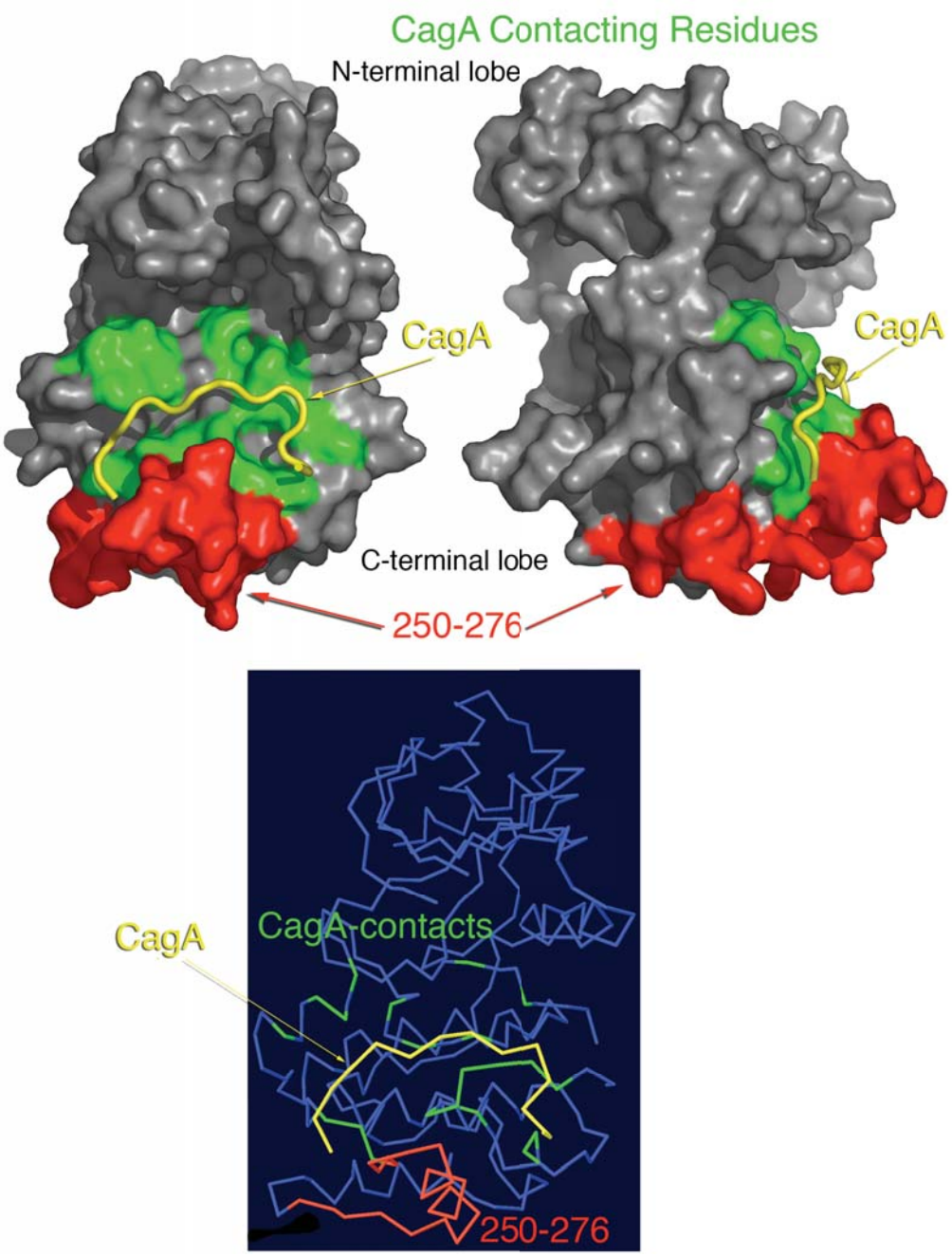
Native electrospray ionization mass spectra of the CagA-MARK2 complex with (a) wild-type CagA, and (b) mutant CagA(LRKL). The electrospray ionization solution had a CagA-MARK2 complex concentration of ~10μM in 300mM ammonium acetate buffer, pH 7.4. The coloured balls above the peaks indicate how many copies of CagA (dark purple) and MARK2 (light orange) are present. Peak assignments were confirmed by performing MS/MS (see Supporting Information). Running conditions and m/z values are summarized in the Supporting Information—Tables 1-3.



Supplementary Fig. 4: MS/MS confirmation of (WT CagA)$_1$(MARK2)$_2$, (WT CagA)$_1$(MARK2)$_1$, and (LKRL mutant CagA)$_1$(MARK2)$_1$ peak assignments in Supplementary Figure 34. (a) Wild-type CagA with isolation of the most abundant CagA:(MARK2)$_2$ ion at ~5250 m/z (17+ charge state). Note that the isolation window also included an adjacent charge state (16+) of this precursor species. The red bar indicates the position of these two precursor ions. Inset shows increased magnification of the high mass region of the MS/MS spectrum. (b) Wild-type CagA with isolation of the most abundant CagA:MARK2 ion at ~3950 m/z (charge state 13+). Note that the isolation window also included an adjacent charge state (12+) of this precursor species. (c) LKRL mutant CagA with isolation of the most abundant CagA:MARK2 ion at ~3925 m/z (charge state 13+). The electrospray ionization solution: 20μM of CagA + MARK2 in 300mM ammonium acetate buffer, pH 7.4.



Supplementary Fig. 5: Effect of increasing concentration of CagA+MARK2 complex on the native mass spectra.



Supplementary Fig. 6: CagA binds MARK2 in the substrate-binding region, near the interface between the two lobes of the kinase (green surface indicate interacting residues), and not in the region 250-276 previously identified through deletion analysis (shown in red). The CagA peptide is shown in yellow. The two upper images show the molecular surface of the kinase, the lower image the trace of the polypeptide. It is likely that the deletion of the large region 250-276 results in a substantial destabilization of the kinase, including those structural elements required to maintain the fold of the substrate binding region where CagA binds.

	Crystal 1 name
Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	93.47, 93.25, 113.95
α, β, γ (°)	90.00, 100.94, 90.00
Resolution (Å)	50-2.20 (2.28-2.20)
R_{sym} or R_{merge}	8.9 (70.6)
$I / \sigma I$	18.9 (2.62)
Completeness (%)	99.9 (100.0)
Redundancy	6.2 (6.2)
Refinement	
Resolution (Å)	41-2.20
No. reflections	97,422
R_{work} / R_{free}	20.4/24.8
No. atoms	
Protein	10,625
Water	687
B-factors	
Protein	25.82
Water	31.32
R.m.s. deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.507

Supplementary Table 1: Data collection and refinement statistics (molecular replacement)

		Mutant			Wild-type		
Species	*Calculated MW*	*z*	*m/z*	*Measured MW*	*z*	*m/z*	*Measured MW*
CagA	*Mutant* 12 751 *Wild-type* 13 034	8+ 7+ 6+ 5+	1641.0 1875.3 2187.4	13 119 ±19	11+ 10+ 9+ 8+ 7+ 6+	1223.2 1343.4 1495.0 1676.3 1915.6 2238.4	13 424 ±17
MARK2	37 536.4	13+ 12+ 11+ 10+	2920.3 3154.5 3440.9 3784.8	37 855 ±21	13+ 12+ 11+ 10+	2911.6 3154.1 3440.7 3784.7	37 837.4 ±0.3
$(CagA)_1$ $(MARK2)_1$	*Mutant* 50 288 *Wild-type* 50 570	14+ 13+ 12+ 11+	3653.4 3931.6 4259.4 4654.7	51 112 ±14	14+ 13+ 12+ 11+	3663.5 3943.4 4271.3 4661.3	51 256 ±14
$(CagA)_1$ $(MARK2)_2$	88 106				20+ 19+ 18+ 17+ 16+ 15+	4452.7 4690.8 4952.2 5241.8 5570.4 5939.6	89 093 ±30

Supplementary Table 2: Summary of peaks from MS spectra. Calculated MW refers to the molecular weight calculated from the protein sequence. Masslynx v4.1 was used to find the m/z values reported here. Measurement errors are the standard deviations of the individual mass determinations made from the different charge states of the given species.

Species	Calculated MW	Mutant parent peak = 3921 z	m/z	Measured MW	Wild-type parent peak = 3945 z	m/z	Measured MW	Wild-type parent peak = 5247 z	m/z	Measured MW
CagA	*Mutant* 12 751	8+	1643.7		8+	1679.2		8+	1676.5	
		7+	1878.3		7+	1913.0		7+	1915.8	
		6+	2187.9	13 135 ±10	6+	2234.9	13 407 ±8	6+	2234.9	13 403.8 ±0.3
CagA*	*Wild-type* 13 033.7				5+	2681.7		5+	2681.8	
					4+	3352.0				
					3+	4468.9				
MARK2	37 536.4							12+	3156.9	
				37 840.9 ±0.7			37 870.4 ±0.1	11+	3443.9	37 871.3 ±0.2
		10+	3785.2		10+	3788.0		10+	3788.1	
		9+	4205.6		9+	4208.8				
MARK2*		7+	5406.9	37 841.4 ±0.6	7+	5411.2	37 871.2 ±0.1			
		6+	6307.9		6+	6312.9				
		5+	7569.4		5+	7575.2				
CagA: MARK2	*Mutant* 50 287.6	13+	3921.1		13+	3944.8				
		12+	4247.9		12+	4274.0				
				50 962.3 ±0.5			51 273 ±3			51 280 ±5
CagA: MARK2*	*Wild-type* 50 570							7+	7326.6	
								6+	8548.8	
								5+	10255.8	
CagA: $(MARK2)_2$	88 107							17+	5247.0	89 180 ±2
								16+	5574.7	
$(MARK)_2$*	75 073							10+	7575.4	75 753 ±12
								9+	8419.9	
								8+	9469	

Supplementary Table 3: Summary of peaks from MS/MS spectra. Calculated MW refers to the molecular weight calculated from the protein sequence. Masslynx v4.1 was used to find the m/z values reported here. Species listed with a * indicate charge states that appear "charge-stripped" (see Supplementary Fig. 4). Measurement errors are the standard deviations of the individual mass determinations made from the different charge states of the given species.

	MS wild-type (Fig. 3)	MS Mutant (Fig. 3)	MS/MS wild-type (Fig. S4)	MS/MS Mutant (Fig. S4)
Capillary voltage (kV)	1.5	1.6	1.5	1.5
Sampling Cone	110.0	89.0	100.0	108.0
Extractor Cone	2.5	1.4	2.5	3.5
Source Temperature (°C)	60	60	60	60
Trap CE (V)	40.0	27.8	120	100
Transfer CE (V)	10.0	6.1	8.4	4.2
Nanoflow pressure (bar)	0.1	0.1	0	0
Backing pressure (mbar)	5.44	5.39	5.32	5.36
Trap pressure (10^{-2} mbar)	2.34	2.14	1.99	1.92
IMS pressure (mbar)	0.52	0.48	0.50	0.41
TOF pressure (10^{-6} mbar)	1.49	1.42	1.19	1.36
# of spectra averaged	300	300	300	300
Smoothing factor	7	7	7	7

Supplementary Table 4: Summary of key instrumental parameters (Waters Synapt nomenclature) used to obtain spectra.