Identification of the periplasmic DNA receptor for natural transformation of *Helicobacter pylori*

Prashant P. Damke¹ ⁵, Anne Marie Di Guilmi¹, Paloma Fernández Varela², Christophe Velours², Stéphanie Marsin², Xavier Veaute³, Mérick Machouri¹, Gaurav V. Gunjal⁴, Desirazu N. Rao⁴, Jean-Baptiste Charbonnier² & J. Pablo Radicella¹*

Horizontal gene transfer through natural transformation is a major driver of antibiotic resistance spreading in many pathogenic bacterial species. In the case of Gram-negative bacteria, and in particular of *Helicobacter pylori*, the mechanisms underlying the handling of the incoming DNA within the periplasm are poorly understood. Here we identify the protein ComH as the periplasmic receptor for the transforming DNA during natural transformation in *H. pylori*. ComH is a DNA-binding protein required for the import of DNA into the periplasm. Its C-terminal domain displays strong affinity for double-stranded DNA and is sufficient for the accumulation of DNA in the periplasm, but not for DNA internalisation into the cytoplasm. The N-terminal region of the protein allows the interaction of ComH with a periplasmic domain of the inner-membrane channel ComEC, which is known to mediate the translocation of DNA into the cytoplasm. Our results indicate that ComH is involved in the import of DNA into the periplasm and its delivery to the inner membrane translocator ComEC.

¹ Institute of Cellular and Molecular Radiobiology, Institut de Biologie François Jacob, CEA, Université de Paris and Université Paris Sud, F-92265 Fontenay aux Roses, France. ² Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, F-91198 Gif-sur-Yvette, France. ³ Institute of Cellular and Molecular Radiobiology, Institut de Biologie François Jacob, CEA, INSERM, Universités Paris Diderot and Paris Sud, F-92265 Fontenay aux Roses, France. ⁴ Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India. ⁵ Present address: Gluck Equine Research Center, University of Kentucky, Lexington, KY 40546, USA. *email: pablo.radicella@cea.fr
**Results**

**ComH is essential for the import of transforming DNA.** *comH* was originally identified as a gene required for NT by screening a mutant library of *H. pylori* for non-transformable strains.6 We confirmed the effect of *comH* inactivation on NT by determining the frequencies of integration of a streptomycin resistant (StrepR) marker using as donor total genomic DNA from a StrepR isogenic strain. When compared to the wild-type strain, the Δ*comH* mutant displayed >10,000-fold reduction in the yield of recombinant clones, an effect similar to that resulting from inactivation of *comB2* or *comEC* (Table 1), both essential for natural competence.\\n
**Table 1 Natural transformation frequencies for *H. pylori* strains.**

| Genotype                  | Mean transformation frequencies | n     | Relative value | P-value (MWU) |
|---------------------------|---------------------------------|-------|----------------|---------------|
| wild-type                 | 1.60 × 10^{-3} (1.87 × 10^{-3}) | 27    | 1.00           |               |
| Δ*comH*                   | 7.07 × 10^{-8} (1.26 × 10^{-7}) | 11    | 4.42 × 10^{-5} | <.0001        |
| Δ*comH rdxA::comH-FLAG*   | 5.54 × 10^{-4} (4.93 × 10^{-4}) | 9     | 0.35           | 0.0356        |
| Δ*comB2*                  | 6.68 × 10^{-4} (4.34 × 10^{-4}) | 4     | 4.17 × 10^{-6} | <.0001        |
| Δ*comEC*                  | 2.45 × 10^{-4} (4.90 × 10^{-4}) | 4     | 1.52 × 10^{-6} | <.0001        |
| Δ*dprA*                   | 3.97 × 10^{-6} (1.69 × 10^{-6}) | 4     | 2.48 × 10^{-3} | <.0001        |
| Δ*recA*                   | 4.05 × 10^{-9} (2.32 × 10^{-9}) | 3     | 2.53 × 10^{-6} | <.0005        |

The recombination frequencies of the isogenic streptomycin resistant (StrepR) total genomic DNA were calculated as the number of streptomycin resistance colonies per recipient colony-forming unit. Values correspond to the mean and standard deviation. n No. of independent determinants. MWU Mann-Whitney U test.
ComH is required for the import of tDNA into the periplasm. During NT in H. pylori, fluorescently labelled tDNA can be transiently detected in the periplasm as discreet foci in wild-type strains but not in those mutated in comB,comH. To better define the role of ComH, we determined the efficiency of tDNA foci formation (Fig. 1a). While ~20% of the wild-type bacteria displayed fluorescent tDNA foci, <1% of the ΔcomB2 cells had detectable foci (Fig. 1b). In a ΔcomEC strain, the proportion of bacteria with tDNA foci was similar to that of the wild-type (Fig. 1b), consistently with the role of ComEC in the DNA transport through the inner membrane. When ComH was disabled, only 1% of the cells presented tDNA foci. Wild-type levels were recovered in the ΔcomH strain by ectopic expression of ComH-FLAG (Fig. 1a, b).

We also monitored by PCR the presence of internalised DNA in the different mutant strains (Fig. 1c, d). After incubation of the cells in the presence of a 6.3 kb linear dsDNA with no homology to the H. pylori genome, the extracellular and unprotected DNA was removed by extensive washing and nuclelease treatment. After isolating total DNA, a 1.1 kb region of the internalised donor dsDNA was amplified by PCR. In agreement with the formation of tDNA foci, amplification of internalised DNA was observed in the wild-type while only residual amplification was detected in a

| Genotype | Mean recombination frequencies | n | Relative value | P-value (MWU) |
|----------|-------------------------------|---|----------------|----------------|
| wild-type | $5.28 \times 10^{-7}$ (5.86 $\times 10^{-7}$) | 16 | 1.00 | |
| ΔcomH | $3.25 \times 10^{-7}$ (1.83 $\times 10^{-7}$) | 5 | 0.61 | 0.9555 |
| ΔcomB2 | $1.90 \times 10^{-6}$ (1.48 $\times 10^{-6}$) | 4 | 3.59 | 0.1482 |
| ΔcomEC | $4.36 \times 10^{-7}$ (5.95 $\times 10^{-7}$) | 9 | 0.82 | 0.4605 |
| ΔdprA | $1.54 \times 10^{-8}$ (2.14 $\times 10^{-8}$) | 5 | 0.029 | <0.001 |
| ΔrecA | 0 | 4 | 0 | 0.0004 |

Exponentially growing H. pylori cells were electroporated with 139-mer single-stranded DNA coding for streptomycin resistance and the recombination frequencies were calculated as the number of streptomycin resistance colonies per recipient colony-forming units. Values correspond to the mean and standard deviation. n No. of independent determinants. MWU Mann-Whitney U test.
ΔcomB2 strain (Fig. 1c, d). Because the level of amplified tDNA in a ΔcomEC strain was similar to that of the wild-type, it can be assumed that it essentially reflects the DNA present in the periplasm (Fig. 1c, d). Disruption of comH led to a dramatic reduction in the levels of internalised DNA, comparable to those of the ΔcomB2 strain, that could be reverted by the ectopic expression of the gene in the ΔcomH rdxA::comH-FLAG strain (Fig. 1c, d). The residual amplification observed in the competent mutant strains is likely due to non-specific DNA binding to the cell surface. Collectively, the foci analysis and the PCR amplification experiments showed that ComH is essential for the presence of tDNA in the periplasm during natural transformation.

ComH is in the periplasm where it co-localises with tDNA foci.

We then proceeded to determine the subcellular localisation of ComH by fractionation of exponentially growing H. pylori ΔcomH cells expressing a FLAG-tagged version of the protein. Western blot analysis showed that the majority of ComH-FLAG was present in the periplasm (Fig. 2a) consistently with the presence of a signal peptide (amino acids 1–19) (http://www.cbs.dtu.dk/services/SignalP/)32 potentially targeting the protein to this compartment. With the fractionation conditions used in this experiment no association of ComH with the membrane fraction was detected. The presence of ComH in the periplasm was confirmed using a strain expressing a HA-tagged version of the protein from the comH locus (Supplementary Fig. 1).

In view of its localisation and its role in the import of tDNA into the periplasm, we asked whether ComH co-localised with the tDNA during natural transformation. To address this question, we generated a H. pylori strain expressing a ComH-mCherry fusion protein yielding transformation frequencies comparable to those of the wild-type strain (Supplementary Table 1). The expression and stability of ComH-mCherry in H. pylori was verified by Western blot (Supplementary Fig. 2). Confocal microscopy analysis revealed that the majority of the cells showed a rather homogenous and weak fluorescent signal on which interspaced puncta or clusters could be observed (Fig. 2b). This pattern was observed independently of the genetic background. The number of clusters increased with time, reaching a maximum 90 min after putting the cells in contact with DNA. When a comEC disrupted strain (comH-mCherry ΔcomEC) was used for the co-localisation experiment in order to stabilise the tDNA foci, we found that 423 out of the 553 tDNA foci analysed (76%) co-localised with ComH-mCherry clusters (Fig. 2b). This was confirmed by a high Manders’ coefficient (0.88 ± 0.02) reflecting the fraction of green signal (DNA foci) with the clusters of red signal (ComH-mCherry) determined using a threshold defined by the JACOP plugin from ImageJ33. The detection of ComH clusters without DNA foci could reflect the presence of non-labelled DNA arising from bacterial lysis. This is consistent with the fact that genetic markers can be transferred from one strain to another by simply co-culturing them.

ComH binds double-stranded DNA. Despite the fact that no obvious DNA binding motif could be identified by sequence analysis of the protein, the co-localisation of ComH with tDNA and the inability of ΔcomH strain to promote tDNA entry led us to hypothesise that ComH could be the periplasmic receptor for the incoming DNA. While in other naturally transformable Gram-negative bacteria this is the role of ComE(A)56, no orthologue of such proteins has been identified in H. pylori. We therefore tested the in vitro DNA binding ability of purified ComH. We first assessed the DNA binding capacity of the purified C-terminally His-tagged fusion protein (ComH-His6) (Supplementary Fig. 3a) by electrophoretic mobility shift assays (EMSA). ComH-His6 formed discrete nucleoprotein complexes with an 18 bp dsDNA in a concentration dependent manner while no binding to ssDNA (18-mer) was detectable (Fig. 3a). The same preference for dsDNA over ssDNA was observed when Surface Plasmon Resonance (SPR) was used to characterise the DNA interaction properties of the ComH-His6. The binding sensorgrams show that ComH-His6 protein bound dsDNA with high affinity while only residual binding was detected when ssDNA was used as substrate (Fig. 3b). Neither the Langmuir isotherm nor the equilibrium approach using the Proteon Manager software gave satisfactory fittings for the calculation of affinity constants. In order to obtain quantitative data on the binding affinity of ComH for dsDNA, we used the microscale thermophoresis (MST) using a 18 bp dsDNA labelled with a 5’FAM34. The observed Kd was 0.05 ± 0.03 µM (Fig. 3c). We also analysed the interaction between ComH and 18 bp dsDNA or 18nt ssDNA using sedimentation velocity analytical ultracentrifugation (SV-AUC). With the dsDNA we observed the formation of a major complex with a sedimentation coefficient (6.5S) and with a Kd of 0.1 ± 0.2 µM (Fig. 3d and Supplementary Fig. 4a). No complex was observed in the same conditions with the ssDNA (Supplementary Fig. 4b).

To determine the DNA binding characteristics of ComH we used the purified MBP-ComH protein (Supplementary Fig. 3b) in EMSA. MBP-ComH showed affinity for a variety of DNA substrates, as long as they harboured a double-stranded region.
ultracentrifugation (SV-AUC) of the interaction between ComH and 18 bp dsDNA-FAM. The free dsDNA-FAM has a sedimentation coefficient of 2.1S. Taken together, the results obtained using two different tagged constructs of ComH protein and four different experimental approaches, show that ComH is a novel DNA binding protein, with strong affinity for duplex DNA compared to ssDNA.

ComH binds to dsDNA through its C-terminal domain. The full length ComH protein is composed of 479 amino acids (54.78 kDa). Preliminary bioinformatics analysis of ComH predicted the presence of two structured regions at the N- and C-termini of the protein connected by a linker sequence. Neither region showed potential structural homology to other proteins and therefore they represent novel and uncharacterised domains of unknown function. To search for the ComH DNA-binding domain, we separately purified the N-terminal (ComH-NTD) (amino acids 20–169) and C-terminal domains (ComH-CTD) (amino acids 170–479) of the protein fused to either His6 or MBP tags (Supplementary Fig. 3a, b). We first tested the DNA binding characteristics of ComH-CTD-His6 and ComH-NTD-His6 using EMSA. As the full-length protein, ComH-CTD-His6 bound dsDNA while it had essentially no affinity for ssDNA (Fig. 4a, b). On the other hand, ComH-NTD-His6 had basically no DNA binding capacity to either dsDNA or ssDNA in the same protein concentration range (Fig. 4b, c). Similar results were obtained when we tested the DNA binding capacities of the MBP fused domains. Both, full-length MBP-ComH and MBP-ComH-CTD formed discrete nucleo-protein complexes when incubated in the presence of dsDNA while the MBP-ComH-NTD displayed negligible DNA binding affinity (Supplementary Fig. 6a, b). In order to obtain quantitative data on the relative binding affinity of full-length ComH versus the N- and C-terminal domains for dsDNA, the interactions of the His6-tagged domains were analysed by MST. The Kd obtained for ComH-CTD was 3.4 ± 1.7 µM (Fig. 4d). While this value reflects a lower affinity than the full-length protein, the ComH-NTD had a much weaker affinity (Kd > 40 µM) (Supplementary Fig. 6c) in agreement with the EMSA analyses. Finally, we addressed the question of whether ComH needed free DNA ends for loading using His6 tagged constructs. Both, ComH-His6 and ComH-CTD-His6 displayed concentration dependent binding to linear and covalently closed supercoiled plasmid DNA irrespective of their topology (Supplementary Fig. 7a, b).

In conclusion, the DNA binding properties of the truncated version of ComH suggest that the CTD harbours a novel DNA binding motif responsible for the affinity of ComH for dsDNA.

ComH-CTD is sufficient to allow tDNA entry into the periplasm. Since the CTD of ComH is sufficient for DNA binding, we tested whether expression of this domain in a ΔcomH strain was sufficient to revert the mutant phenotype. For this purpose, as we did for the full length protein, we inserted in rdxΔ the...
ComH-CTD-FLAG coding sequences fused to the comH upstream sequences and the N-terminal signal peptide from ComH (Fig. 5a). We then confirmed that ComH-CTD-FLAG was expressed and exported to the bacterial periplasm like the full-length form (Fig. 5b). The ectopic expression of ComH-CTD-FLAG resulted in a partial restoration of tDNA foci formation (Fig. 5c). The capacity of the CTD to allow the import of the tDNA to the periplasm was confirmed by the duplex PCR assay. Indeed, ectopic expression of ComH-CTD in the ΔcomH strain allowed significant amplification of the tDNA retained by the bacteria after nuclease treatment, albeit to lower levels than those observed for the mutant strain complemented with the full-length ComH (Fig. 5d).

We then assessed the capacity of the truncated form of ComH to allow the whole natural transformation process up to the chromosomal integration of the tDNA. Unlike that of the full-length ComH-FLAG, the ectopic expression of ComH-CTD-FLAG in the ΔcomH strain did not restore the capacity to obtain recombinants (Fig. 5e).

Taken together, these results suggest that while the CTD of ComH is sufficient for the normal detection of the tDNA in the periplasm of competent cells during NT, it does not support its translocation into the cytoplasm. These results were further confirmed following the internalisation of fluorescently labelled DNA into the bacterial cytoplasm, as previously described11 (Supplementary Movies 1–3). While in a wild-type strain the volume of internalised DNA increased over time, in a ΔcomEC mutant the tDNA remained blocked in the periplasm, reflecting the impairment in this mutant of the transport through the inner membrane. When the ΔcomH rdxA::comH-CTD-FLAG strain was tested in this assay, it displayed a phenotype very similar to that of the ΔcomEC mutant (Supplementary Fig. 8).

ComH NTD interacts with the periplasmic OB domain of ComEC. The similitude of phenotypes between the ΔcomEC and ΔcomH rdxA::comH-CTD-FLAG strains (periplasmic foci formation but no translocation of the tDNA into the cytoplasm) raised the possibility that ComH might act in close association with ComEC to allow the transport of tDNA through the inner membrane. Topological prediction of H. pylori ComEC structure36 suggests that this inner membrane protein has 9 transmembrane segments and an oligonucleotide binding (OB) domain present on the periplasmic side of the membrane (Fig. 6a). We therefore tested a potential interaction between ComH and the ComEC-OB domain. Extracts from E. coli bacteria expressing His6-tagged recombinant fusions of the ComEC-OB domain were incubated with E. coli extracts containing either MBP or MBP-ComH. The products of amylose affinity pull-down from the mixtures were then eluted from the resin by competition with maltose and probed by immunoblot for the presence of the ComEC-OB domain using anti-His tag antibodies. As shown in Fig. 6b, ComEC-OB-His6 was recovered from the pull-down when co-incubated with full-length MBP-ComH, indicating a direct interaction between ComH and the OB domain from ComEC. To rule out the possibility that the interaction was mediated by DNA present in the extracts, the pull-down experiments were repeated with the addition of benzonase. The interaction between ComH and the ComEC OB domain persisted even after degradation of the DNA (Supplementary Fig. 9a, b). Interestingly, the His6-tagged OB domain was also recovered when MBP-ComH-NTD was used as bait but not in the case of MBP-ComH-CTD or MBP alone. These results showed that through its NTD, ComH is able to interact with ComEC, providing a link between the handling of tDNA in the periplasm and its passage through the inner membrane.
Fig. 5 Ectopic expression of ComH DNA binding domain (ComH-CTD) restores DNA uptake but not transformation. a Schematic representation of ectopically expressed FLAG-tagged ComH constructs in a ΔcomH genetic background. The upstream sequence (~200 bp) and the coding sequence for the signal peptide of comH (SP) were added to maintain the expression levels of the gene close to that of the wild-type strain. b Subcellular localisation of ComH-FLAG and ComH-CTD-FLAG was monitored by immunoblotting using anti-FLAG antibodies. T: Total extract, P: periplasm, C: cytoplasm. c Fluorescent DNA (ATTO-550-dUTP labelled 408 bp dsDNA) foci formation in the ΔcomH rdxA::comH-CTD-FLAG strain. 3394 cells were counted from at least two independent experiments. 256 (7.5%) bacteria showed presence of detectable DNA foci. Scale bar = 2.5 μm. d Detection of internalised donor DNA in the ΔcomH rdxA::comH-CTD-FLAG strain by the PCR assay. e Comparison of recombination frequencies after natural transformation with genomic DNA of wild-type and ΔcomH rdxA::comH-CTD-FLAG strains. The central line represents the median, the bounds of the boxes show the inner quartile range and the whiskers represent the minimum and maximum values of the data set. The numbers above the box refer to P-values calculated using nonparametric MWU T tests.

Discussion

During natural transformation, H. pylori displays a remarkable efficiency to import tDNA into the periplasm of the bacteria and its subsequent recombination in the genome. Pioneering work in H. pylori showed that this pathogen uses a two-step DNA uptake mechanism in which first the dsDNA is transported from the surface of the cell to the periplasm and subsequently it is transferred to the cytoplasm through the inner membrane. Similar two-step DNA uptake mechanisms were described for the Gram-negative bacteria N. gonorrhoeae and V. cholerae. While the first step is mediated by species-specific DNA uptake machineries, the ComB type IV secretion system in the case of H. pylori or type IV (pseudo-) pili for the other naturally transformable bacteria, the second step of DNA translocation across the cytoplasmic membrane is mediated by the conserved ComEC (ComA in Neisseria) channel. In N. gonorrhoeae and V. cholerae, it is proposed that the periplasmic DNA binding protein ComEA (ComE in Neisseria) is required for DNA traversal of the outer membrane by retaining it within the periplasm.

In the absence of a comEA orthologue in H. pylori, we searched for proteins able to bind the transforming DNA in the periplasm. Here, we showed that ComH is the H. pylori periplasmic DNA receptor. The comH gene (hp1527) is critical for natural transformation, the transformation frequencies of the comH null mutants being close to the detection limit (Table 1). Amino acid sequence analysis of ComH indicated that it is unique to H. pylori with apparently no homology to known protein motifs or domains, making it difficult to speculate on its function during competence. The only recognisable feature of the sequence is a
putative signal peptide that would target the protein to the periplasm. Despite complete sequence dissimilarity to ComEA proteins, we found that ComH shows the essential characteristics of periplasmic DNA receptor proteins. Indeed, ComH is present in the periplasm, binds dsDNA with high affinity and is required for the formation of tDNA foci, all properties shared with ComEA. While most of the ComEA proteins harbour conserved structural helix-hairpin-helix (HhH) motifs crucial for the DNA binding and thus contains a novel DNA binding fold. Further structural studies should allow its definition. As it is the case for ComEA with the Type IV pili involved in the DNA uptake in other species, whether ComH interacts with the ComB components remains to be determined.

How the DNA is handed to the ComEC inner-membrane channel is still unknown. It has been proposed that the presence in the periplasmic space of ComEC loops or domains would allow the interaction of the membrane-embedded channel with other DNA-uptake machinery proteins. Interestingly, the capacity of ComH CTD to bind dsDNA is sufficient for the import of the incoming DNA into the periplasm but not for its translocation into the cytoplasm, indicating a separation of functions. Transection of the NTD leads to the persistence of DNA foci in the periplasm, a phenotype observed in comEC mutants. These results suggest that this domain is implicated in the delivery of the DNA to the inner membrane pore. Our observation of a physical interaction between ComH NTD and the OB domain from ComEC, predicted to be exposed to the periplasm, is consistent with a role for ComH in the handing of the DNA to the inner membrane ComEC channel. It would be interesting to investigate if in the case of other naturally transformable bacteria ComEA is also implicated in the delivery of the tDNA to ComEC (ComA in Neisseria).

Based on the results presented here, we propose that ComH represents a new class of DNA binding proteins and is not only the missing DNA receptor for natural transformation in H. pylori but also the protein responsible for delivering the DNA to ComEC. Its role in the translocation of tDNA through the bacterial envelope can be divided in two steps as represented in Fig. 7. By its capacity for binding ssDNA through its CTD, ComH acts as a receptor for the incoming DNA into the periplasm (Fig. 7a). Subsequently, ComH could deliver the DNA to the inner-membrane channel through the interaction of the NTD with ComEC OB domain (Fig. 7b). Although we have here unveiled the role of a new player in the mechanism of transforming DNA internalisation in H. pylori, several questions remain open. In S. pneumoniae, the nuclease EndA is required to generate the single stranded DNA capable of crossing the inner-membrane but a protein with similar function remains to be identified in other species. It has been proposed that ComEC could harbour such an activity. Furthermore, the need of other yet unknown proteins to process the dsDNA or to coordinate the different steps of the DNA passage through the periplasm cannot be ruled out.

**Methods**

**H. pylori growth conditions.** H. pylori cultures were grown under microaerophilic conditions (5% O2, 10% CO2, using the MAC-MIC system from AES Chemunex) at 37°C. Blood agar base medium (BAB) supplemented with 10% defibrillated horse blood (AES) was used for plate cultures. Liquid culture were grown in brain heart infusion media (BHI) supplemented with 10% defibrillated and de-complemented fetal bovine serum (Invitrogen, Carlsbad, CA, USA) with constant shaking (180 rpm). Antibiotic mix containing polymyxin B (0.155 mg/ml), vancomycin (6.25 mg/ml), trimethoprim (3.125 mg/ml), and amphotericin B (1.25 mg/ml) was added to both plate and liquid cultures.

**Construction of gene variants.** H. pylori 26695, a kind gift from Agnès Labigne (Institut Pasteur), was used as parental strain to generate all the gene variants listed in Supplementary Table 2. The gene sequences were obtained from the annotated complete genome sequence of H. pylori 26695 deposited at Pylori Gene World Wide Web Server (http://genolist.pasteur.fr/PyloriGene/). For chromosomal integrations/replacements of gene variants, the specific gene regions (with 200–300 bp of flanking sequences wherever appropriate) were amplified from genomic DNA (gDNA) of H. pylori 26695 strain using sequence specific primers (Supplementary Table 3) and cloned in pJET1.2 (2974 bp) vector by blunt end ligation. Further, the cassette encoding either the antibiotic resistance or indicated protein tag were introduced in the gene construct by either classical restriction-ligation method or using sequence- and ligation-independent cloning (SLIC). All the plasmids harbouring different gene constructs (listed in Supplementary Table 4) were verified by DNA sequencing and introduced into H. pylori by natural transformation. Transformants were selected by using appropriate required antibiotics: kanamycin (20 μg/ml), apramycin (12.5 μg/ml), and chloramphenicol (8 μg/ml). The correct constructions were verified by PCR using locus and gene specific primers. The glycerol stocks of H. pylori were prepared in BHI media supplemented with 12.5% glycerol and stored in -80 °C. The details of different constructions generated in this study are given below.
Construction of comH::Km/Cm. lpcomH (hp1527) locus was amplified using primers Op81 and Op82 and ligated to blunt pET1.2 vector to generate p1196 (pET1.2-comH[Km][Sup]). The sequence encoding the plasmid p1196 (using primers Op83 and Op84), non-polar kanamycin (using primers Kpn1-Km-for and BamH1-Km-rev), and chloramphenicol (using primers Kpn1-Cm-for and BamH1-Cm-rev) resistance cassette were digested using Kpn1 and BamH1 and ligated separately to generate p1198 (pET1.2-comH[Cm][Sup]) and p1295 (pET1.2-comH[Cm][Cm]). p1198 and p1295 were used for allelic replacement of native comH locus.

Ectopic expression of comH-FLAG and ComH-CTD-FLAG. hp1527 locus (+200 bp upstream sequence) was amplified using Op99 and Op100 and ligated to nDXA-Cm cassette present in the plasmid p1081 (amplified using primers Op101 and Op102) using SLIC to generate plasmid p1235 (pJet1.2-Prom-comH-FLAG). The gene sequence coding for ComH-NTD was deleted by reverse PCR of p1235 using Op200 and Op202 to generate plasmid p1337 (pJet1.2-Prom-comH-CTD) (158-479-FLAG-Cm). p1235 and p1337 were used to transform wild-type strain followed by disruption of the native comH locus using p1198. To ectopically express ComH-FLAG in GFP expressing bacteria, the chloramphenicol cassette of p1337 was replaced by apramycin cassette (plasmid p1457) and the plasmid was used to transform strain LR887 (268695 pUrea-GFPmut2-Km) followed by disruption of native comH locus using p1295.

Constructions of translational fusions. Natural transformations assays were performed as described in. Briefly, exponentially growing H. pylori cells (optical density of 4.0 at 600 nm) were incubated with isogenic streptomycin resistance total chromosomal DNA (200 ng) on Blood Agar Base (BAB) plates for 24 h at 37 °C. Next, day, serial dilutions of H. pylori cells were plated on BAB plates with or without streptomycin (10 μg/ml) and incubated for 4–5 days at 37 °C. The recombinant colonies were calculated using a calibrated Abbe nomocolor to determine the resistance colonies per recipient colony-forming unit. P values were calculated using the Mann–Whitney U test on GraphPad Prism software.

Determination of spontaneous mutation frequencies. Exponentially growing H. pylori cells (optical density of 0.4 at 600 nm) were spotted on Blood Agar Base (BAB) plates for 24 h at 37 °C (no exogenous DNA was added). Next day, the spots were collected in peptone water. 200 μl of the decimalised sample was directly plated on BAB plates with or without streptomycin (10 μg/ml) and incubated for 4–5 days at 37 °C. The recombinant colonies were calculated using a calibrated Abbe nomocolor to determine the resistance colonies per recipient colony-forming unit.

Determination of recombinant fusions after electroporation. Chemically competent electro-competent cells were prepared by re-suspending exponentially growing H. pylori cells (optical density of 10 OD/ml at 600 nm) in 1 ml of ice-cold 15% Glycerol, 9% Sucrose. After centrifugation (2204 g for 3 min), the bacterial pellet was re-suspended in 250 ml of ice-cold 15% Glycerol, 9% Sucrose. 50 μl of these electro-competent cells were mixed with 1 μl of salt free sDNA in a clean electroporation cuvette. The electroporation was performed at 2.5 kV cm–1, and 25 μF.

Subcellular fractionation. To obtain the subcellular fractions, 20 ml of bacterial cultures at OD600 1.5 were pelleted, washed in PBS and resuspended in 2 ml buffer C (20% Sucrose, 1 mM EDTA, 30 mM Tris-HCl, pH8) and further incubated for 10 min at room temperature and centrifuged at 5700 × g. The pellet was resuspended in 1 ml of ice-cold water and incubated 10 min at 4 °C. After centrifugation at 5,766 g, the supernatant corresponding to the periplasmic fraction was recovered. The pellet was resuspended in 1 ml of buffer A (10 mM Tris–HCl, pH 7.5, 1 mM DTT, 1x protease inhibitor cocktail) and sonicated. After centrifugation at 142,470 × g, the cytoplasmic fraction (supernatant) and membrane (pellet resuspended in 0.2 ml of 30 mM Tris pH 8, 100 μM NaCl) fractions were recovered. The presence of proteins in different bacterial compartments was monitored by immunoblotting.

Uncovered images are provided as Source Data. As controls for the fractionation experiments, we determined the localisation of MotB and NdrK, membrane and cytoplasmic proteins, respectively.

Duplex PCR assay. The presence of transforming DNA in the periplasm of H. pylori was detected using protocol adapted from ref. Briefly, Linear plasmid DNA (6.3 kb) was prepared by restriction digestion of p1391 by Ncol-HF. Exponentially growing H. pylori cells (1 OD/ml at 600 nm) were incubated with 250 ng of plasmid DNA in BHI supplemented with 10% fetal bovine serum for 15 min at 37 °C under microaerophilic conditions. The cells harvested by centrifugation were washed twice with 1x PBS. The excess non-transformed DNA was degraded by nuclease cocktail [DNasel (50 U/ml), Benzonase (62.5 U/ml)] in 1x PBS, and 10 mM MgCl2 treatment at 37 °C for 20 min. The cells were further washed twice, and re-suspended in 200 μl of 1x PBS. This resuspension was subjected to phenol-chloroform extraction and the supernatant was recovered. 1 μl of supernatant was directly used as template for PCR using QS DNA polymerase. A 1.1 kb donor specific region was amplified using primers Op284 and Op285, while a 1.4 kb H. pylori gDNA specific region (hp0247) was amplified using primers Op31 and Op32. The PCR fragments were visualised by agarose gel electrophoresis (1%) stained with CBB-Gold (1:10,000 dilution). Uncropped gel images are provided as Source Data.

Fluorescently labelled DNA substrates for microscopy. DNA substrates were prepared as described in. Briefly, 408 bp ATT-Oligo DNA was prepared by amplification of hp1197 locus from 26695 gDNA (100 ng) using primers 1197–5 and 1197–3 (0.5 μM each), 250 μl of dNTP mix, 5 μ of ExTaq enzyme (Takara), supplemented with 10 μg of ATT-500-aminolyl-duTTP (Jena bioscience). 2 kb ATT-488-duTTP labelled dsDNA was prepared by amplification of hp1527 locus from 26695 gDNA (100 ng) using primers pMALp2X (amplified by DNA polymerase A. 1.1 kb donor specific region was amplified using primers Op284 and Op285, while a 1.4 kb H. pylori gDNA specific region (hp0247) was amplified using primers Op31 and Op32. The PCR fragments were puriﬁed by Illustra GFX puriﬁcation kit (GE Healthcare Little Chalfont, UK).

Fluorescence microscopy experiments. Both live and ﬁxed H. pylori cells were prepared as described. Briefly, 408 bp ATT-Oligo DNA was prepared by amplification of hp1197 locus from 26695 gDNA (100 ng) using primers 1197–5 and 1197–3 (0.5 μM each), 250 μl of dNTP mix, 5 μ of ExTaq enzyme (Takara), supplemented with 10 μg of ATT-500-aminolyl-duTTP (Jena bioscience). 2 kb of ATT-488-duTTP labelled dsDNA was prepared by amplification of hp1527 locus from 26695 gDNA (100 ng) using primers pMALp2X (amplified by DNA polymerase A. 1.1 kb donor specific region was amplified using primers Op284 and Op285, while a 1.4 kb H. pylori gDNA specific region (hp0247) was amplified using primers Op31 and Op32. The PCR fragments were puriﬁed by Illustra GFX puriﬁcation kit (GE Healthcare Little Chalfont, UK).

Cloning, over-expression and puriﬁcation of ComH and constructs. Escherichia coli strains used for cloning (DH5α), and for protein overexpression, and puriﬁcation (BL21-DE3 [Bacteria–Bertani] were cultured in LB broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) with 1x antibiotics. The bacterial strain followed by disruption of the native comH locus by electroporation experiment. The sequence encoding amino acids 20–479 of ComH (amplified using Op220 and Op221) was ligated to pET-AH–vH (ampliﬁed using primers Op219 and Op226) using SLIC to generate plasmid p1364 [pET-AH–vH (comH (20–479)]. His-tagged fusion constructs of ComH-NTD (amino acids 20–169) and ComH-CTD (amino acids 170–479) were generated by reverse PCR of p1364 using primer pairs (Op239 and Op240), and (Op243 and Op244) respectively. To purify the MBP fused constructs under the control of IPTG inducible promoter, the sequence encoding amino acids 20–479 of ComH (amplified using Op127 and Op82) was ligated to pMAL-p2X (amplified using primers Op128 and Op129) using SLIC to generate p1249 [pMAL-p2X –MBP-comH (20–479)]. MBP tagged fusion constructs of ComH-NTD (amino acids 20–169) and ComH-CTD (amino acids 170–479) were generated by reverse PCR of p1249 using primer pairs (Op198 and Op205), and (Op206 and Op201) respectively. All the plasmids containing fusion constructs were verified by DNA sequencing and are listed in Supplementary Table 4.
For large-scale purification, the plasmids carrying different recombinant DNA constructs were transformed into E. coli BL21 expression strain and grown in LB media containing ampicillin (100 µg/ml) at 37 °C. Induction of the proteins were achieved by IPTG (0.5 mM) at an optical density of 0.6 to 0.65 nm for 4–5 h at 30 °C. The cells harvested by centrifugation were resuspended in lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 0.2% NP40, and 1% protease inhibitor cocktail). 5 mM imidazole was added in case of the Hisn-tagged proteins. After sonication, the soluble fractions were separated from the cell debris by centrifugation at 18,637 × g for 45 min. Standard Ni2+-NTA, and Amylose affinity chromatography was used for the first step of purification of the proteins. The fractions containing the proteins were pooled, and desalted using HiTrap desalting columns (GE Healthcare). Full-length ComH, and ComH-NTD were further purified using Resource-S, while the ComH-CTD was further purified using Resource-Q ion-exchange chromatography. Purity of the proteins was assessed by SDS-PAGE and the pure fractions were pooled and dialysed twice for 4–6 h against storage buffer (50 mM Tris, 100 mM NaCl, 1 mM DTT, 0.1 mM EDTA, and 10% glycerol). The proteins were quantitated using Bradford assay and small aliquots were frozen immediately in liquid N2 and stored at –80 °C.

**Electrophoretic mobility shift assay**. HPLC purified, chemically synthesised oligonucleotides (5′-Cy5)-3′-TAMRA labelled) were used for the EMSA experiments (Supplementary Table 5). Different DNA substrates were prepared by annealing appropriate combinations of oligonucleotides (Supplementary Table 5). Indicated concentrations of purified proteins were incubated with DNA substrates (30 nM) in binding buffer (10 mM Tris-HCl pH 7.5, 50 mM KCl, 1 mM DTT, 0.1 µg/µl BSA). The NaCl concentration was maintained at constant amount (~100 mM) by adding protein storage buffer to the reaction mixture. After incubation of 30 min on ice, the reaction mix was separated by native-PAGE (6%). The gels were visualised using by Typhoon and quantified using ImageJ. Percentage binding was estimated by measuring the depletion in substrate DNA considering DNA without protein as 100%.

**Surface plasmon resonance**. SPR binding kinetics were analysed using the BIACore3000 optical biosensor (GE Healthcare) at 25 °C. A 30-mer 5′-biotinylated DNA (a-fam) was immobilised on a streptavidin-coated surface biosensor chip. The double-stranded was generated by hybridising the complementary DNA strand (Op250). Variable concentrations of ComH-His6 (0.05% v/v) were further purified using Resource-S and a high sedimentation complex with sedimentation coefficient 3.8S and a high sedimentation complex with sedimentation coefficients of 6.5S and 6.5S complexes are 0.1 and 0.3 µM and may correspond to ComH-dsDNA and [ComH]2-dsDNA, respectively.

**Sedimentation velocity analytical ultracentrifugation (SV-AUC)**. The interactions between ComH-His6 and dsDNA or ssDNA were measured using an Optima XL-A ultracentrifuge (Beckman Coulter). Samples were centrifuged at 48,000 × g for 5 min and the absorbance at 280 nm was measured. The sedimentation coefficients were determined using the SEDNTERP program. The sedimentation velocity coefficients of 6.5S and 6.5S complexes are 0.1 and 0.3 µM and may correspond to ComH-dsDNA and [ComH]2-dsDNA, respectively.

**MBP pull-down assay**. The coding sequence (amino acids 30–156) of hpcmEC amplified using primers HpcmEC-30_F and HpcmEC-156_R2 was ligated to plasmid pMBE (digested with Ndel – BamHI) using SfIc to generate plasmid p1397. The interaction between MBP-ComH and ComEC-OB-His6, was tested by amylose affinity pull-down assay. The plasmids carrying these recombinant constructs were transformed into E. coli BL21 cells. The expression of the proteins was induced by IPTG (0.5 mM) and cultures were grown for 3 h at 37 °C. Equal amounts of these E. coli cultures were mixed together and sonicated. After centrifugation, the soluble fraction of the cell lysate was incubated with amylose resin for overnight at 4 °C to allow the interaction between the proteins. The resin was washed, and the bound proteins were eluted using maltose (10 mM). Equal amounts of the elution fractions were separated by SDS-PAGE. The presence of ComEC-OB-His6 in the eluted fractions was detected by Western blots using anti-His antibodies. Uncropped gel images are provided as Source Data.

**Data availability**. The authors declare that the main data supporting the findings of this study are available within the article and its Supplementary information files. The source data underlying Figs 1b–d, 2a, 3a, c, 4a–b, d are provided as Source Data. Extra data are available from the corresponding author upon request.

Received: 11 April 2019; Accepted: 31 October 2019; Published online: 25 November 2019

**References**

1. Burton, B. & Dubnau, D. Membrane-associated DNA transport machines. Cold Spring Harb. Perspect. Biol. 2, a004006 (2010).
2. Johnston, C., Martin, B., Fichant, G., Poland, P. & Claverys, J. P. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat. Rev. Microbiol. 12, 181–196 (2014).
3. Kruger, N. J. & Stingl, K. Two steps away from novelty: principles of bacterial DNA uptake. Mol. Microbiol. 80, 860–867 (2011).
4. Stingl, K., Muller, S., Scheidgen-Kleyboldt, G., Clausen, M. & Maier, B. Composite system mediates two-step DNA uptake into Helicobacter pylori. Proc. Natl Acad. Sci. USA 107, 1184–1189 (2010).
5. Görgöl, G., et al. Concerted Intra- and Extracellular Dynamics of Imported DNA and ComEC DNA Uptake Protein during Concomitant Transformation. PLoS Pathog. 10, e1004043 (2014).
6. Seitz, P. et al. ComEC Is Essential for the Transfer of External DNA into the Periplasm in Naturally Transformable Vibrio cholerae Cells. PLoS Genet. 10, e1004066 (2014).
7. Seitz, P. & Blokesch, M. DNA transport across the outer and inner membranes of naturally transformable Vibrio cholerae is spatially but not temporally coupled. MBio 5, 4 (2014).
8. Chen, I., Christie, P. J. & Dubnau, D. The ins and outs of DNA transfer in bacteria. Science 310, 1456–1460 (2005).
9. Pelicic, V. Type IV pili: e pluribus unum. Mol. Microbiol. 82, 837–838 (2007).
10. Hepp, C. & Maier, B. Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism. Proc. Natl Acad. Sci. USA 113, 12467–12472 (2016).
11. Salzer, R., Kern, T., Joos, F. & Averhoff, B. The Thermus thermophilus comEA/comEC operon is associated with DNA binding and regulation of the DNA translocator and type IV pili. Environ. Microbiol. 18, 65–74 (2016).
12. Takeno, M., Taguchi, H. & Akamatsu, T. Role of ComEC in DNA uptake during transformation of competent Bacillus subtilis. J. Bacteriology-Biofilm. 113, 689–693 (1929).
13. Draskovic, I. & Dubnau, D. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulfide bonds. Mol. Microbiol. 55, 881–896 (2005).
14. Backert, S., Niedermann, M., Maubach, G. & Naumann, M. Pathogenesis of Helicobacter pylori infection. Helicobacter 21(Suppl 1), 19–25 (2016).
15. Suerbaum, S. & Michetti, P. Helicobacter pylori infection. N. Engl. J. Med. 347, 1175–1186 (2002).
16. Oddot, X. et al. Genomic evolution and transmission of Helicobacter pylori in two South African families. Proc. Natl Acad. Sci. USA 110, 13880–13883 (2013).
17. Falush, D. et al. Recombination and mutation during long-term gastric colonization by Helicobacter pylori: estimates of clock rates, recombination size, and minimal age. Proc. Natl Acad. Sci. USA 98, 15056–15061 (2001).
18. Gressmann, H. et al. Gain and loss of multiple genes during the evolution of Helicobacter pylori. PLoS Genet. 1, e43 (2005).
Krennemann, L. et al. Helicobacter pylori genome evolution during human infection. *Proc. Natl Acad. Sci. USA* **108**, 5033–5038 (2011).

Kraft, C. & Suerbaum, S. Mutation and recombination in Helicobacter pylori: mechanisms and role in generating strain diversity. *Int. J. Med. Microbiol.* **295**, 299–305 (2005).

Suerbaum, S. & Achtman, M. Evolution of Helicobacter pylori: the role of recombination. *Trends Microbiol.* **7**, 182–184 (1999).

Levinson, D. O., Pfaller, S., Henke, H. G. & Haas, R. Natural competence for DNA transformation in Helicobacter pylori: identification and genetic characterization of the comB locus. *Mol. Microbiol.* **28**, 1027–1038 (1998).

Karnholz, A. et al. Functional and topological characterization of novel components of the comB DNA transformation competence system in *Helicobacter pylori*. *J. Bacteriol.* **188**, 882–893 (2006).

Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori* *H. pylori* 26695. *Proc. Natl Acad. Sci. USA* **99**, 16898–16903 (2002).

Suerbaum, S. & Kraft, C. Mutation and recombination in *Helicobacter pylori*: identification and genetic characterization of the comB locus. *Mol. Microbiol.* **28**, 1027–1038 (1998).

Levinson, D. O., Pfaller, S., Henke, H. G. & Haas, R. Natural competence for DNA transformation in Helicobacter pylori: identification and genetic characterization of the comB locus. *Mol. Microbiol.* **28**, 1027–1038 (1998).

Karnholz, A. et al. Functional and topological characterization of novel components of the comB DNA transformation competence system in *Helicobacter pylori*. *J. Bacteriol.* **188**, 882–893 (2006).

Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori* *H. pylori* 26695. *Proc. Natl Acad. Sci. USA* **99**, 16898–16903 (2002).