Cell Host & Microbe

Supplemental Information

Structural Insight into Polymorphic ABO Glycan Binding by *Helicobacter pylori*

Kristof Moonens, Pär Gideonsson, Suresh Subedi, Jeanna Bugaytsova, Ema Romaõ, Melissa Mendez,, Jenny Nordén Mahsa Fallah, Lena Rakhimova, Anna Shevtsova, Martina Lahmann, Gaetano Castaldo, Kristoffer Brännström, Fanny Coppens, Alvin W. Lo, Tor Ny, Jay V. Solnick, Guy Vandenbussche, Stefan Oscarson, Lennart Hammarström, Anna Arnqvist, Douglas E. Berg, Serge Muyldermans, Thomas Borén, and Han Remaut

Figure S1

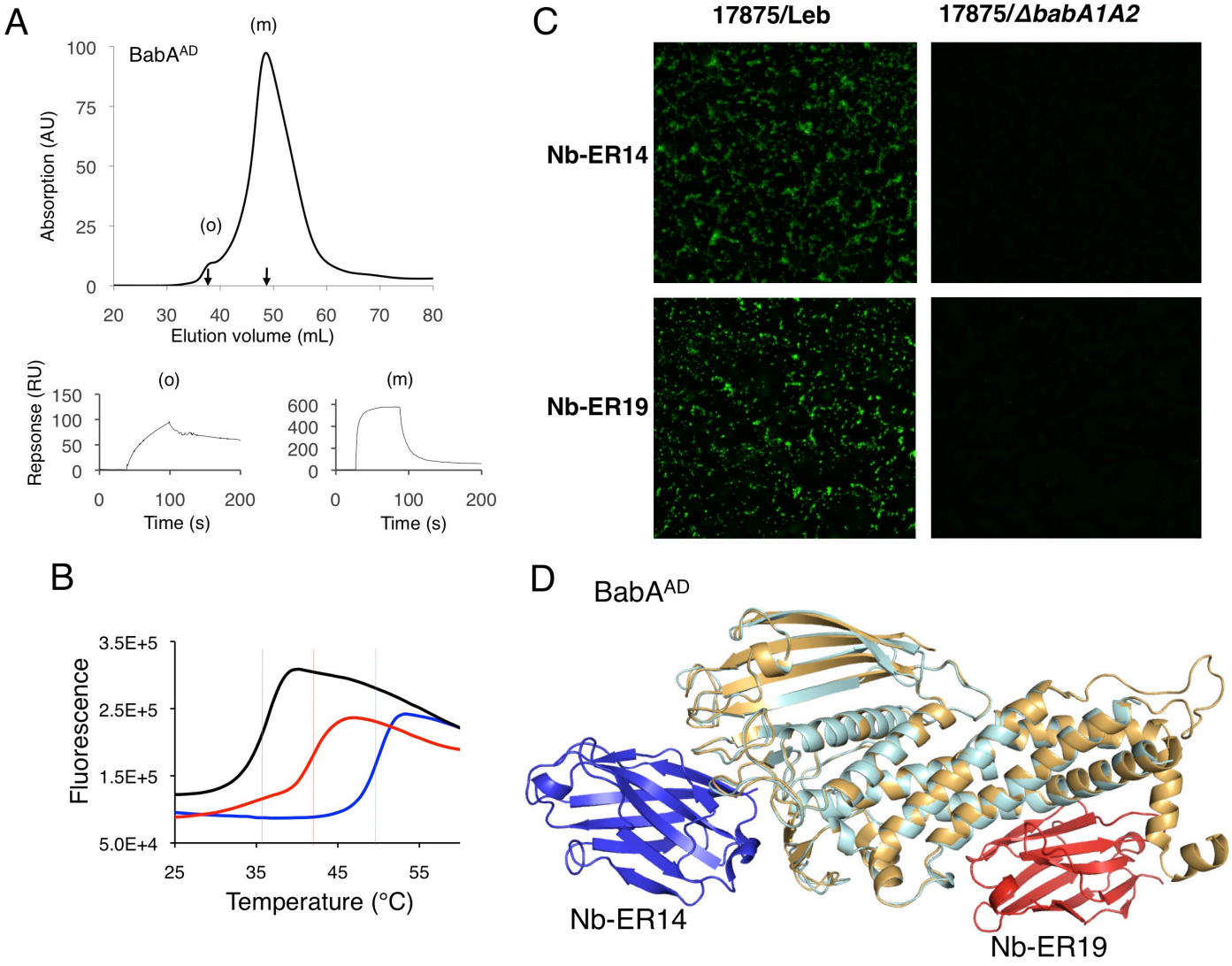

A
BabA^AD
(m)
(o)
Absorption (AU)
Elution volume (mL)
Repsonse (RU)
Time (s)
B
Fluorescence
Temperature (°C)
C
17875/Leb
17875/ΔbabA1A2
Nb-ER14
Nb-ER19
D
BabA^AD
Nb-ER14
Nb-ER19

Figure S2

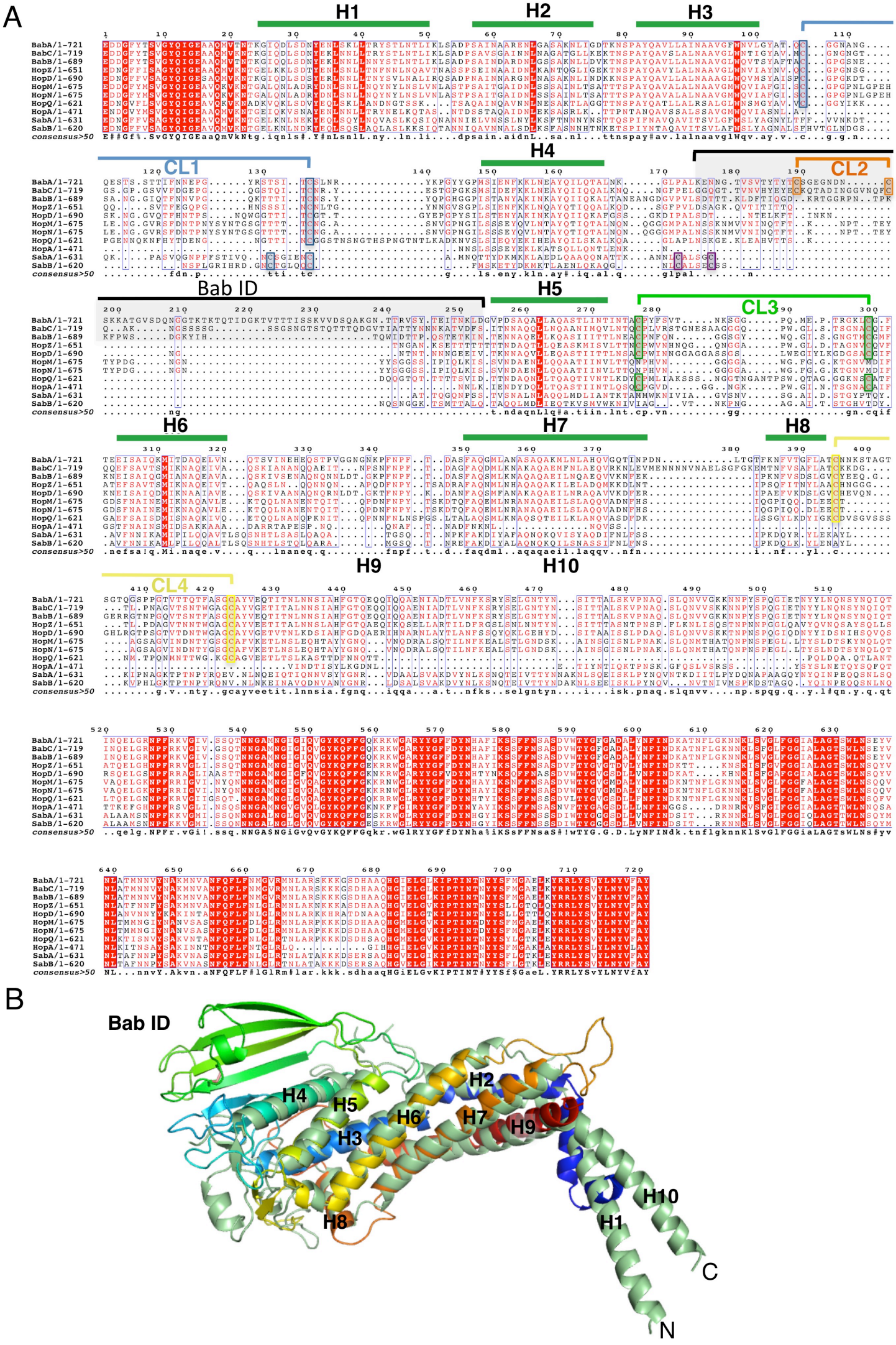

A
H1
H2
H3
H4
CL1
CL2
Bab ID
H5
CL3
H6
H7
H8
CL4
H9
H10
B
Bab ID
H1
H2
H3
H4
H5
H6
H7
H8
H9
H10
C
N

Figure S3

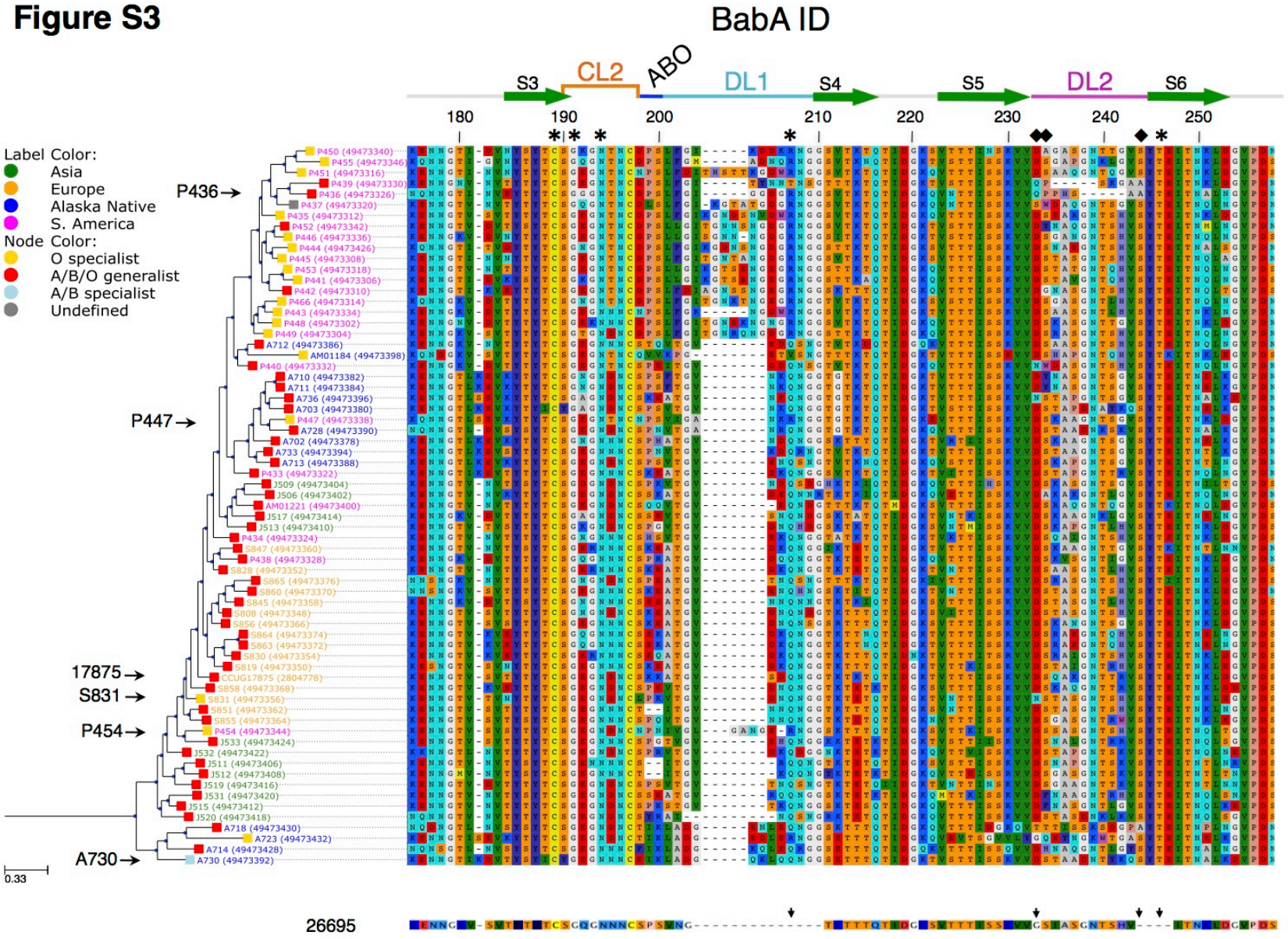

BabA ID
S3
CL2
ABO
DL1
S4
S5
DL2
S6
180
190
200
210
220
230
240
250
Label Color:
Asia
Europe
Alaska Native
S. America
Node Color:
O specialist
A/B/O generalist
A/B specialist
Undefined
P436
P447
17875
S831
P454
A730
0.33
26695

Figure S4

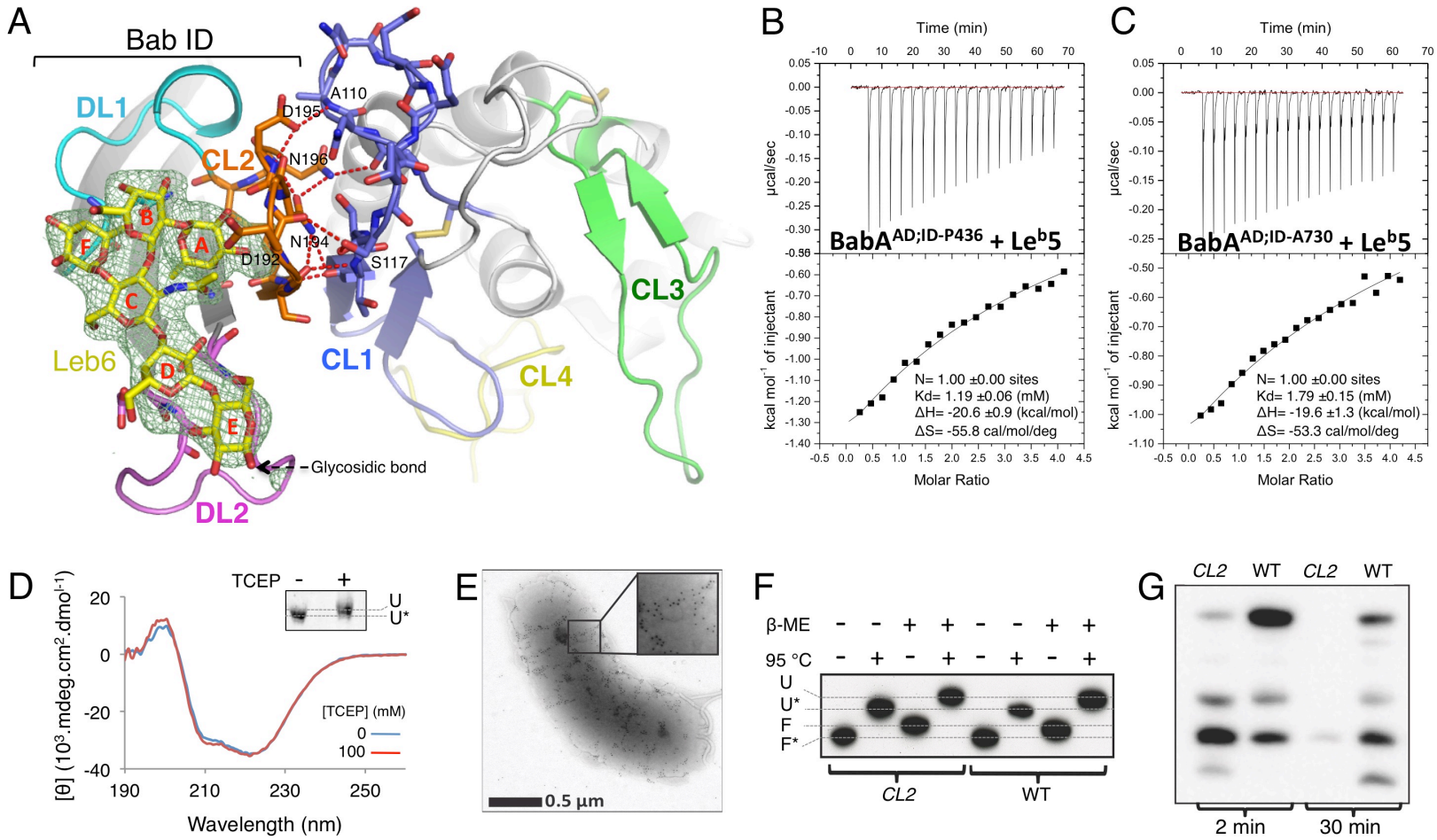

A
Bab ID
DL1
CL2
A110
D195
N196
N194
D192
S117
CL3
CL1
CL4
Leb6
Glycosidic bond
DL2
B
Time (min)
µcal/sec
BabA^AD;ID-P436 + Le^b5
kcal mol^-1 of injectant
N= 1.00 ±0.00 sites
Kd= 1.19 ±0.06 (mM)
ΔH= -20.6 ±0.9 (kcal/mol)
ΔS= -55.8 cal/mol/deg
Molar Ratio
C
Time (min)
µcal/sec
BabA^AD;ID-A730 + Le^b5
kcal mol^-1 of injectant
N= 1.00 ±0.00 sites
Kd= 1.79 ±0.15 (mM)
ΔH= -19.6 ±1.3 (kcal/mol)
ΔS= -53.3 cal/mol/deg
Molar Ratio
D
TCEP - +
U
U*
[θ] (10^3.mdeg.cm^2.dmol^-1)
[TCEP] (mM)
0
100
Wavelength (nm)
E
0.5 µm
F
β-ME - - + + - - + +
95 °C - + - + - + - +
U
U*
F
F*
CL2
WT
G
CL2 WT CL2 WT
2 min
30 min

Figure S5

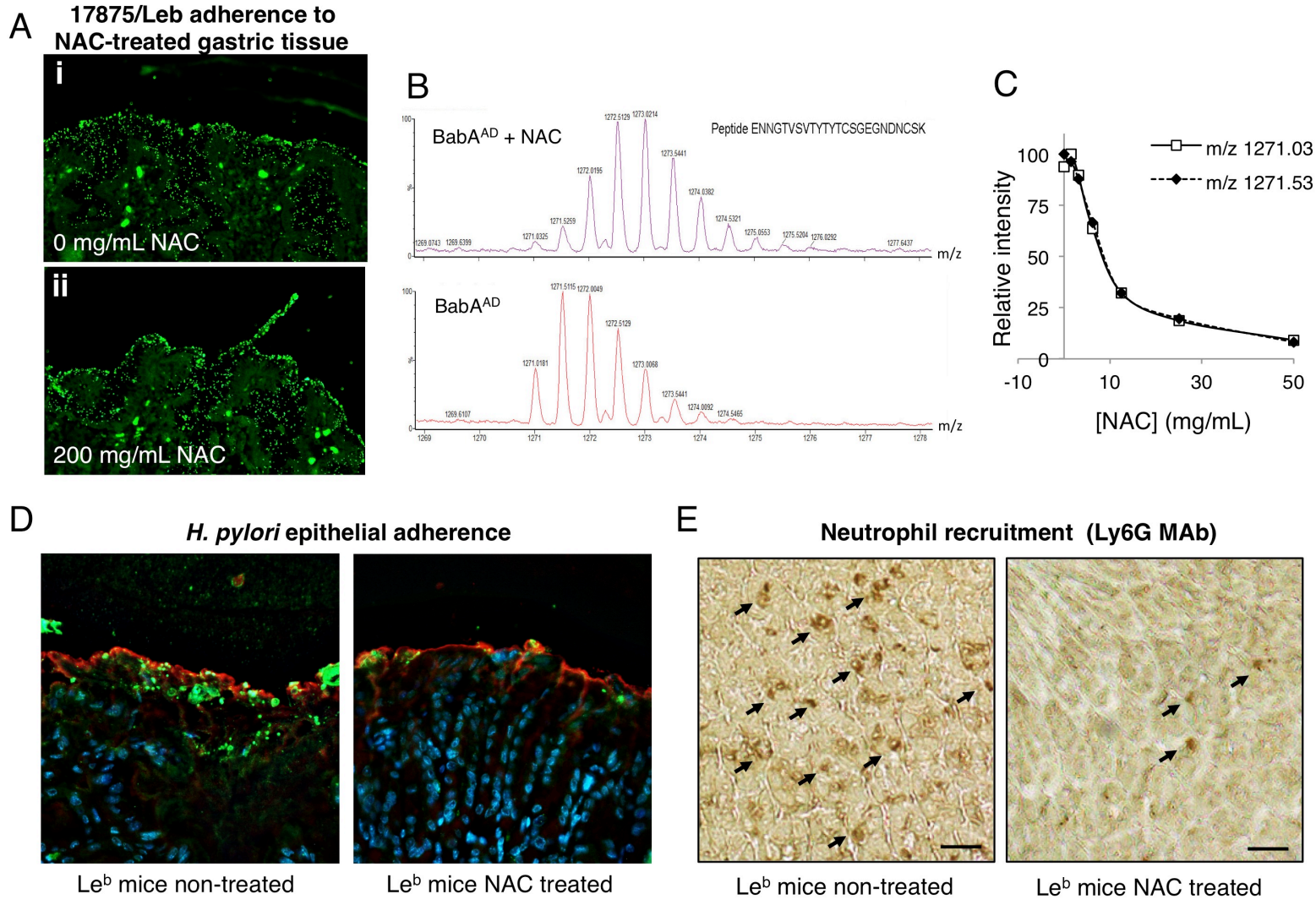

A
17875/Leb adherence to NAC-treated gastric tissue
i
0 mg/mL NAC
ii
200 mg/mL NAC
B
BabA^AD + NAC
Peptide ENNGTVSVTYTYTCSGEGNDNCSK
m/z
BabA^AD
m/z
C
m/z 1271.03
m/z 1271.53
Relative intensity
100
75
50
25
0
-10
10
30
50
[NAC] (mg/mL)
D
H. pylori epithelial adherence
Le^b mice non-treated
Le^b mice NAC treated
E
Neutrophil recruitment (Ly6G MAb)
Le^b mice non-treated
Le^b mice NAC treated

Figure S6

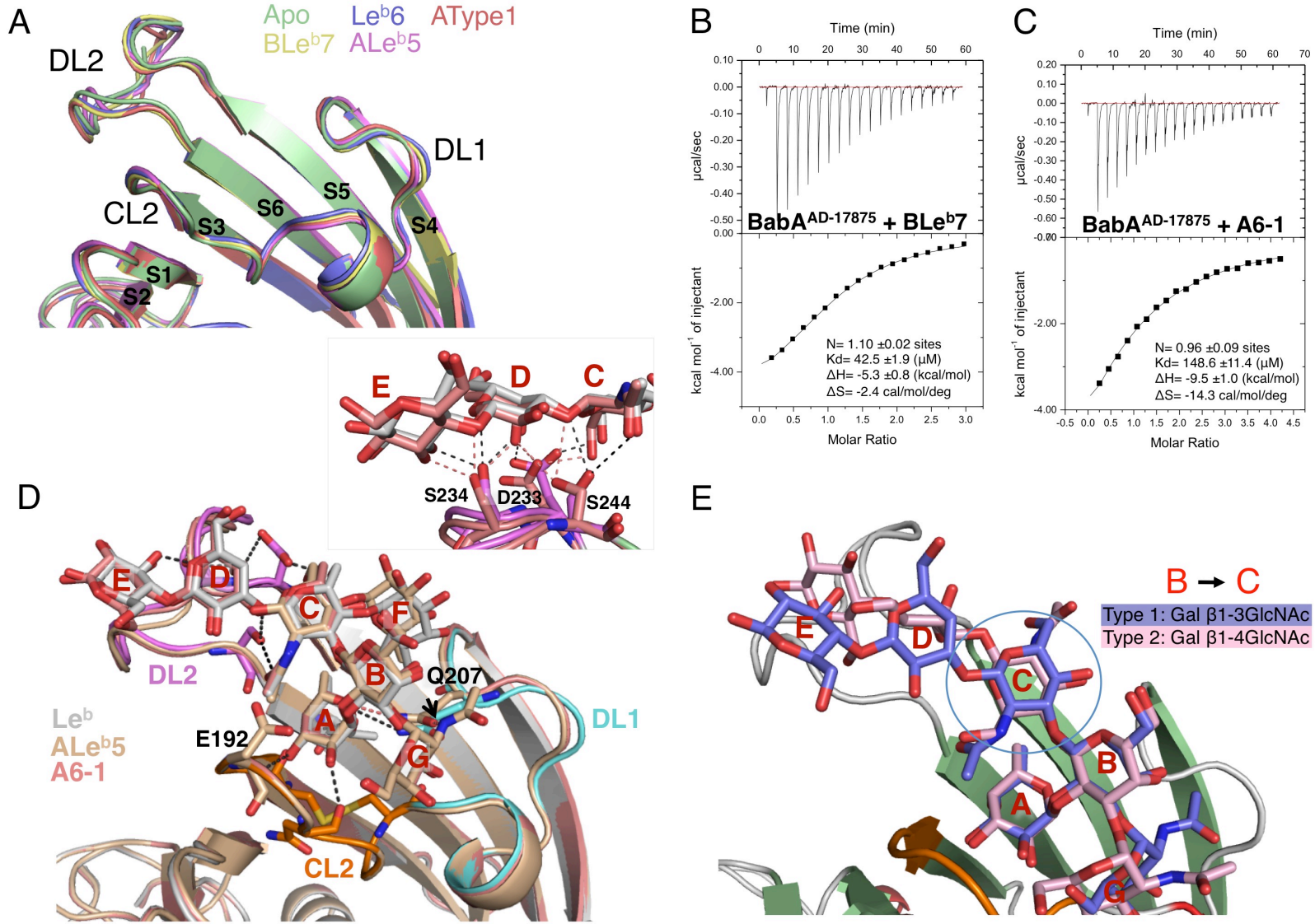
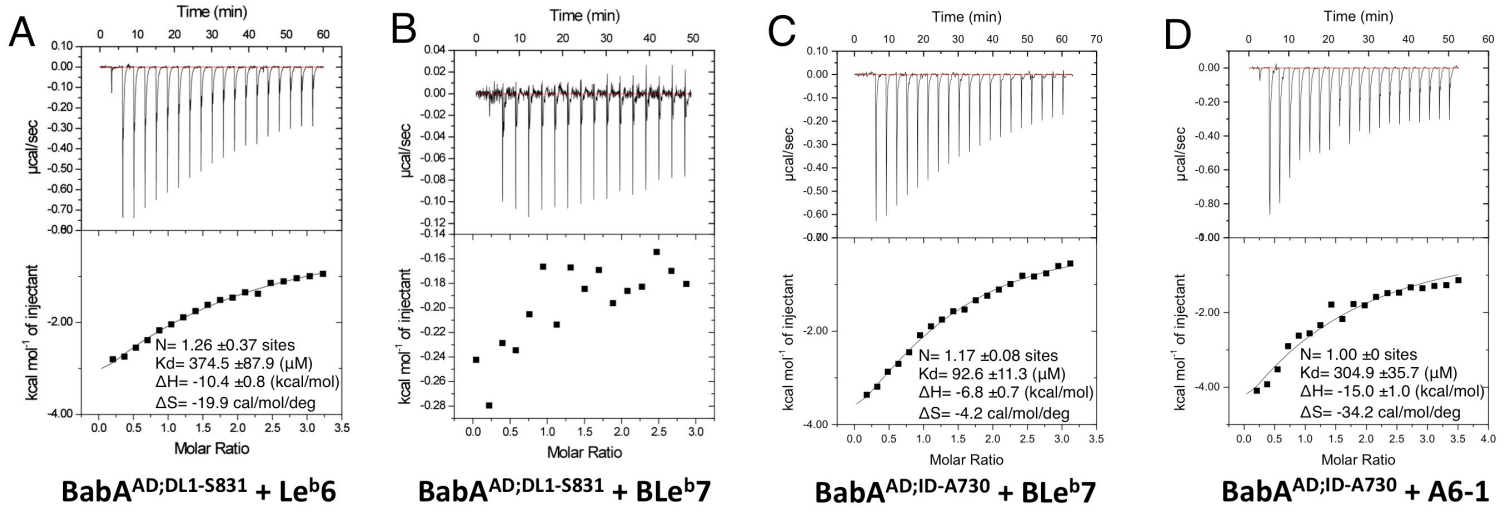
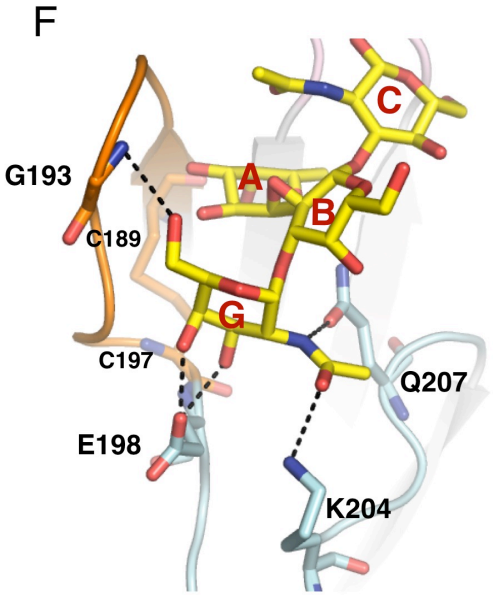
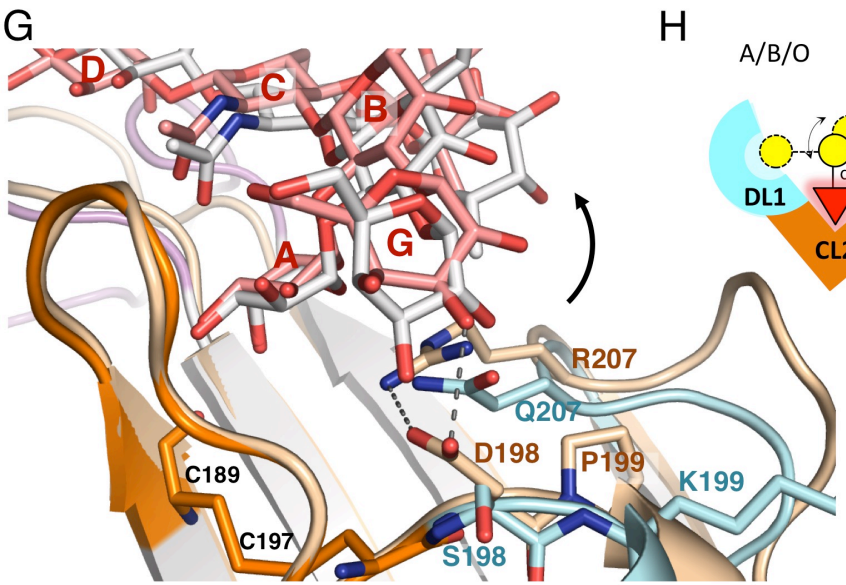
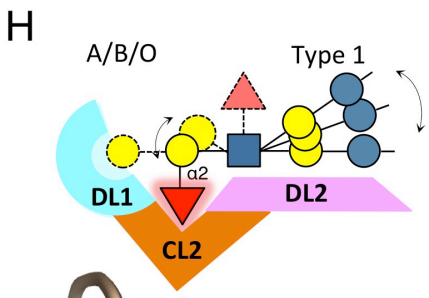

A
Apo
Le^b6
AType1
BLe^b7
ALe^b5
DL2
DL1
CL2
S1
S2
S3
S4
S5
S6
E
D
C
S234
D233
S244
B
Time (min)
µcal/sec
BabA^AD-17875 + BLe^b7
kcal mol^-1 of injectant
N= 1.10 ±0.02 sites
Kd= 42.5 ±1.9 (µM)
ΔH= -5.3 ±0.8 (kcal/mol)
ΔS= -2.4 cal/mol/deg
Molar Ratio
C
Time (min)
µcal/sec
BabA^AD-17875 + A6-1
kcal mol^-1 of injectant
N= 0.96 ±0.09 sites
Kd= 148.6 ±11.4 (µM)
ΔH= -9.5 ±1.0 (kcal/mol)
ΔS= -14.3 cal/mol/deg
Molar Ratio
D
E
D
C
F
DL2
B
Q207
A
G
DL1
Le^b
ALe^b5
A6-1
E192
CL2
E
B → C
Type 1: Gal β1-3GlcNAc
Type 2: Gal β1-4GlcNAc
E
D
C
B
A
G

Figure S7


A
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio
N= 1.26 ±0.37 sites
Kd= 374.5 ±87.9 (µM)
ΔH= -10.4 ±0.8 (kcal/mol)
ΔS= -19.9 cal/mol/deg
BabA AD;DL1-S831 + Le b 6
B
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio
BabA AD;DL1-S831 + BLe b 7
C
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio
N= 1.17 ±0.08 sites
Kd= 92.6 ±11.3 (µM)
ΔH= -6.8 ±0.7 (kcal/mol)
ΔS= -4.2 cal/mol/deg
BabA AD;ID-A730 + BLe b 7
D
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar Ratio
N= 1.00 ±0 sites
Kd= 304.9 ±35.7 (µM)
ΔH= -15.0 ±1.0 (kcal/mol)
ΔS= -34.2 cal/mol/deg
BabA AD;ID-A730 + A6-1


F
G193
C189
C197
E198
K204
Q207
A
B
C
G
BabA AD;ID-A730 + A6-1


G
D
C
B
A
G
R207
Q207
D198
P199
K199
C189
C197
S198
BabA AD;ID-P436 / BabA AD;ID-17875 + BLe b 7


H
A/B/O
Type 1
α2
DL1
DL2
CL2

SUPPLEMENTARY FIGURES LEGENDS

Figure S1 - related to Figure 1: Nanobodies recognize native BabA and stabilize BabA^{AD}. **(A)** Top: SEC profile of purified BabA^{AD} reveals the presence of an oligomeric (o) and monomeric (m) species. Bottom: SPR sensorgrams of fresh SEC fractions corresponding to oligomeric (o) and monomeric BabA^{AD} (m) (indicated with arrows) reveal a slow and fast dissociating species, respectively. Interstitial fractions show a gradient of bipartite dissociation profiles with increasing contribution of fast versus slow dissociation species reflecting increasing (m)/(o) ratios when moving towards longer elution volumes (not shown). At equilibrium, concentrated BabA^{AD} samples show bipartite SPR dissociation profiles reflecting a mixed (o) and (m) population (Figure 1B). **(B)** Thermal unfolding curves of BabA^{AD} (black), BabA^{AD}:Nb-ER14 (blue) and BabA^{AD}:Nb-ER19 (red) as monitored by SYPRO orange fluorescence in function of temperature. Nb-ER14 and Nb-ER19 increase melting mid temperature (Tm) with 13 and 6 °C, respectively. **(C)** *In vitro* adherence assay of FITC labeled nanobodies Nb-ER14 and Nb-ER19 to *H. pylori* strain 17875 and an isogenic *babA1/babA2* knock-out. Both nanobodies efficiently and specifically recognize functional full length BabA expressed on *H. pylori* cells. **(D)** Overlay of the BabA^{AD}:Nb-ER14 (lightblue:blue) and BabA^{AD}:Nb-ER19 (gold:red) structures in ribbon representation. Both Nb-ER14 and Nb-ER19 bind regions with increased thermal displacement factors and local disorder, indicating zones with increased conformational dynamics. The role of either nanobody in facilitating crystallization is by protein stabilization (see (B)), by reduction of local conformational flexibility, and by providing additional crystal contacts.

Figure S2 - related to Figure 2: Domain architecture and secondary structure conservation within the HOP family. (A) Multiple sequence alignment of members of the HOP family members with a predicted secondary structure topology homologous to that found in BabA. Conserved α-helices of the BabA core domain are indicated by green bars over the sequence and named H1 to H10 as in Figure 1. **(B)** The BabA (colored according to chain) and SabA adhesin adhesin domains (pale green; PDB: 4O5J) superimpose with a root mean squared deviation (RMSD) of 2.3 Å for 240 equivalent Ca atoms. Conserved helices in the HOP family are named and the N- and C- termini of the SabA adhesin domain are indicated.

Figure S3 - related to Figure 3: Sequence polymorphism in the BabA insertion domain. MSA of the BabA ID from various *H. pylori* strains with known generalist / specialist profile and geographic background (Aspholm-Hurtig et al., 2004). The geographical origin of the strains is indicated together with the distribution in specialist (only binding O antigen), generalist (binding A/B/O antigens) or undefined (<1 % Le^b RIA binding) strains. Secondary structure

matching the sequence stretches and the region important for Le^b binding are shown above the alignment (asterisks: α1-2 fucose binding pocket; diamonds: type 1 core binding region). CL2: Cys189/Cys197-loop, DL: Diversification Loop, ABO: specialist/generalist determining region. South American strains are predominantly of specialist phenotype. These strains show the specialist Asn/Asp/Leu-Pro signature in position 198-199, as well as the substitution of Asn207 with an Arg and the frequent occurrence of a ~7 residue insertion in DL1. The role of these latter two sequence signatures is presently unknown. Four strains show an exceptional generalist phenotype despite a specialist Asp-Pro genotype (eg. P436, P439, P442 and P452; according (Aspholm-Hurtig et al., 2004)). The non-canonical generalist binding in these strains is the result from loss-of-function mutations in the DL2 site (see Figure S7F and discussion). Shown underneath the MSA is the BabA ID sequence for *babA* from TIGR reference genome of *H. pylori* 26695. This strain lost Le^b adhesion due to mutations in the secretor fucose and type 1 chain binding sites (indicated with arrows).

Figure S4 –related to Figure 4: Sequence conservation around the bg antigen binding site. **(A)** Overview of the interactions between CL1 in the 3+4-helix bundle domain and the BabA carbohydrate binding domain (color scheme as in Figure 3). Residues interacting with the bg antigens are shown in stick model and labeled; CL1 and CL2 are colored blue and orange, respectively. The side chains of the conserved Asn194 and Asp195 (N/D/T) do not directly interact with the Le^b glycan, but likely provide additional conformational stability to the secretor fucose-binding CL2 by means of H-bonding with CL1. Panel includes electron density from a FoFc omit map contoured at 2.5 σ and displayed around the Le^b6 ligand in pale green mesh. Le^b6 is depicted in stick model and colored yellow, red and blue for respectively carbon, oxygen and nitrogen atoms. The arrow indicates the reducing end that will make the glycosidic bond to the neoglycoconjugate. **(B, C)** ITC injection heats (upper) and normalized binding isotherms (lower) of Le^b5 titrations of 17875 $BabA^{AD}$ mutants in which the carbohydrate binding domain (BabA-ID, Figure S3) is replaced with that of strain P436 (B) or A730 (C). For both BabA isoforms, sequence substitutions in the DL2 region reduce monovalent binding affinities to low micromolar dissociation constants **(D)** Far UV CD spectra of 3 mg/mL $BabA^{AD}$ in absence (blue) and presence (red) of 100 mM TCEP show the adhesin domain retains its secondary structure content upon full reduction of its disulfide bonds. The upward mobility-shift of the TCEP-treated protein in coommasie-stained SDS-PAGE (inset) confirms disulfide reduction (U: unfolded – disulfide reduced, U*: unfolded – disulfide bonded). **(E)** The J166*CL2* mutant BabA protein was readily immuno-gold stained in whole cells, showing it is efficiently exported to the outer membrane. **(F)** Folding and disulfide formation in the J166*CL2* mutant protein was confirmed by western blots of SDS solubilized whole cell protein extracts of WT and mutant strains: their

similar heat-modifiable appearance demonstrates folding of the BabA β-barrel (U: unfolded, F: folded), and faster migration of non-reduced (U*, F*) relative to reduced samples (U, F) indicates a compacted protein due to disulfide formation in CL1, CL3 and CL4. **(G)** The J166*CL2* mutant showed increased protease sensitivity at 2 and 30 min of limited trypsinolysis, albeit with a cleavage pattern identical to that of WT protein. This implies a localized increase in structural flexibility rather than an overall loss of folding and conformation.

Figure S5 – related to Figure 5: NAC treatment reduces bacterial binding by reduction of CL2. (A) FITC labeled 17875/Leb bacteria binding to human gastric tissue sections, after a 2 hour pre-treatment of the sections with 0 (i) or 200 (ii) mg/mL NAC shows NAC pre-treatment does not perturb the integrity of the gastric mucosa or the availability of *H. pylori* bg antigen binding sites. **(B)** Mass spectrometry analysis of trypsinized BabA^{AD} with or without prior exposure to 50 mg/mL N-acetylcysteine. The isotopic mass distribution profile for the double-charged Cys189/Cys197 disulfide containing peptide Glu176-Lys199 is shown. **(C)** Relative intensity of ESI-TOF species with m/z values corresponding an oxidized CL2 loop (fragment 176-199) released after tryptic digestion of BabA^{AD} exposed to an increasing concentration of N-acetylcysteine. **(D)** Example gastric epithelium sections of mice treated for 2 weeks with 0 (-) or 40 (+) mg/day NAC in their drinking water, shown to visualize how *H. pylori* epithelial adherence (Figure 5C) was estimated. Antibody-staining reveals *H. pylori* in green (anti-*H. pylori*), Lewis b in red (anti-Le^b antibodies) and nuclei of mouse tissue cells in blue (DAPI). **(E)** Representative immunohistostained (anti-Ly6G antibodies) gastric sections of mice treated or non-treated with NAC, used to determine neutrophil infiltration. See extended experimental procedures section 7 for details on neutrophil and *H. pylori* infectious load in the epithelial proximity.

Figure S6 – related to Figure 6: BabA^{AD} binding to ABO blood group antigens

(A) Overlay of the CBD's in the X-ray structures of apo BabA^{AD}:Nb-ER19 (green) as well as BabA^{AD}:Nb-ER19 in complex with Le^{b}6 (blue), BLe^{b}7 (yellow), ALe^{b}5 (violet) and the A type 1 hexasaccharide A6-1 (brick), shows the high degree of conformational constraint on CL2. **(B, C)** ITC injection heats (upper) and normalized binding isotherms (lower) of BabA^{AD}:Nb-ER19 titrated with BLe^{b}7 (B) or A6-1 (C). **(D)** Overlay of the BabA^{AD}:Le^{b}6 complex (colored as Figure 3B) and the BabA^{AD} in complex with ALe^{b}5 (Table S1; tan) or A6-1 (Table S1; brick). Side chain representation and H-bond labeling as Figure 6. Arrows highlight bg A specific H-bonding. In the Inset: in the A6-1 complex, the interaction of the type 1 chain core interacts with the Asp-Ser-Ser triad through an alternative H-bonding pattern compared to Le^{b}6 or BLe^{b}7 (not shown), possibly due to a slight reorganization of the type 1 chain core as a result of the lack of the Lewis fucose in

the A6-1 glycan. (**E)** Overlay of the BabA^{AD}:Nb-ER19 structure in complex with A6-1 with an energy minimized model (Glycam) of the A type 2 core hexasaccharide. Type 1 and type 2 core glycans differ in having a β1-3 versus β1-4 Gal-GlcNAc linkage (residue B to C, Table S1). The accompanying reorientation of the type 2 core reducing end breaks up the H-bond network with the Asp233-Ser234-Ser244 triad and creates a steric clash with the BabA DL2 loop.

Figure S7 – related to Figure 7: Exchange of the 17875 DL1 loop with strain S831 sequence alters bg polymorphism (A, B) ITC injection heats (upper) and normalized binding isotherms (lower) of S831 DL1 graft mutant BabA$^{AD;DL1\text{-}S831}$:Nb-ER19 titrated with Le^{b}6 (A), or BLe^{b}7 (B). The S831 loop mutant loses binding towards the B antigen, while still retaining affinity for the H antigen. Compared to the native 17875, the S831 DL1 graft mutant shows a ~4.6 fold lower affinity for Leb6. The thermodynamic signature of the ITC data shows the lowered affinity stems from an increased entropic cost of Le^{b}6 binding to the S831 DL1 region compared to the 17875 parent (ie., -19.9 cal/mol/deg versus -13.8 cal/mol/deg, respectively), suggesting the S831 DL1 region is more conformationally dynamic prior to ligand binding than is the 17875 DL1. **(C, D)** ITC injection heats (upper) and normalized binding isotherms (lower) of the 17875 BabA^{AD} grafted with the strain A730 CBD, titrated with BLe^{b}7 (C) or A6-1 (D). Increased H-bond formation of the A730 DL1 loop with the bg A or bg B GalNAc or Gal (see panel E) results in a ~20 and 6-fold increase in monovalent binding affinities compared to that of Le^b (Figure S4C). Thus A730 shows a bg A/B specialist type Le^b binding. **(E)** Close-up view of the DL1 region of the strain A730 CBD bound to A6-1. The altered sequence in the A730 DL1 results in the loss of the canonical α–helical turn seen in strain 17875, S831 or P436 and results in its straightening into a hairpin loop (see also Figure 3B). The extended DL1 conformation allows improved contact with the bg A Gal by means of 5 H-bond interactions. For easy comparison, the labeled residues are numbered according the equivalent residue position in strain 17875. A730 is part of a small monophyletic cluster of Alaskan native isolates with a common DL1 "IKLADG" sequence in residues 199-204, which are expected to result in a similar extended DL1 conformation (Figure S3). In combination of a weaker binding interaction with the type 1 chain core at the DL2 site (Figure 3B), the increased interaction with the bg determinants at the DL1 site make A730 strain effectively an A/B specialist. **(F)** Structural comparison of strain P436 (tan) and strain 17875 (light grey; CL2, DL1 and DL2 colored orange, cyan and magenta, respectively) interacting with BLe^{b}7 (colored light red and grey in P436 and 17875, respectively). P436 holds the D198-P199 bg specialist genotype typical of South American Amerindian isolates, yet shows a generalist binding phenotype. D198 and R207 (numbered according equivalent position in the 17875 reference structure; Figure S3) are in an electrostatic interaction and are positioned inside the bg A GalNAc or bg B Gal substituent binding pocket that is seen in the 17875 BLe^{b}7 complex (this

figure and Figures 6 and 7A). In P436, loss of function mutations in DL2 (S234P and S244A) disrupt the interaction with the type 1 chain core (see also Figure 3B for the P436-Le^{b}6 complex), which results in a ~ 10° upward rotation of the BLe^{b}7 around the secretor fucose (labeled A). The upward rotation of BLe^{b}7 positions the bg B determining Gal (labeled G) above, rather than in steric clash with the D198-R207 pair, thus leading to a non-canonical generalist bg binding phenotype in this otherwise specialist genotype strain. Non-canonical generalist binding profiles in 3 additional SAA specialist genotype strains (Figure S3; strains P436, P439 and P452) is in each case accompanied with similar loss of function mutations in DL2, suggesting the compensatory rotation of the glycan is a common mechanism to circumvent specialist genotype in these strains. **(G)** Schematic representation of BabA's three-pronged ABO Le^b / Type 1 binding site. The disulfide-clasped CL2 wraps the glycans' secretor fucose and forms the structurally conserve anchor point in the binding site. DL1 and DL2 are high sequence diversity regions that interact with the bg A or B substitutuents or type 1 chain reducing end, respectively. Amongst different clinical isolates, sequence variation in these regions result in various extends to which DL1 and DL2 contribute to the glycan interaction and provides a basis to the high divergence in ABO/Le^b glycan binding affinities observed amongst isolates and populations (Aspholm-Hurtig et al., 2004). Variation in the DL1 region results in the observed bg preferences of the *H. pylori* isolates (see main text). Mutations in DL2 determine the extend to which the Le^b type 1 chain core contributes to the binding interaction, tuning the binding affinity of the glycan, but also allowing compensatory reorientation of the glycan with respect to the DL1 bg preference determining site as seen in non-canonical generalist strain P437 (Figure S7F) or the bg A/B specialism as seen in Alaskan strain A730 (Figure S7E). Apart from altering the adhesins' binding properties, sequence diversity in DL1 and DL2 may create antigenic variation for the escape of neutralizing antibodies.

Table S1: glycans used in this study, Figures 1-7

Le^{b}5
Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ1-3Gal

Le^{b}6
Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glc

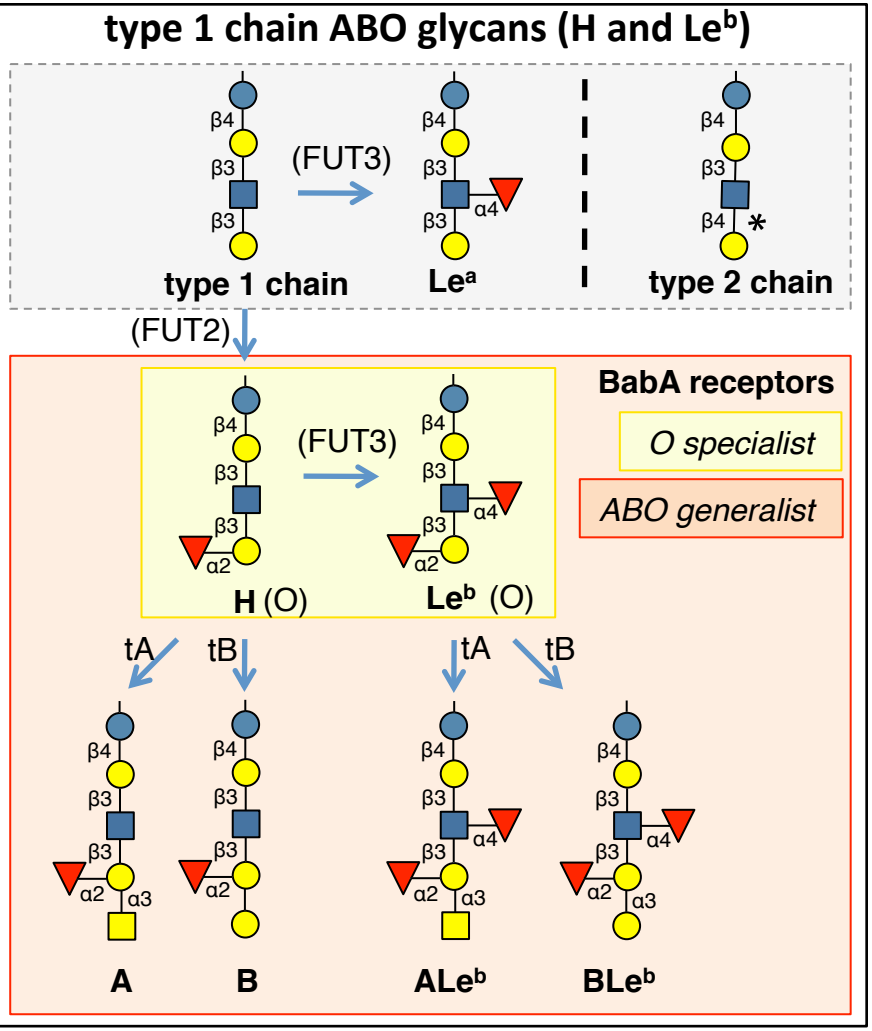
BLe^{b}7
Galα1-3(Fucα1-2)Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glc

ALe^{b}5
GalNAcα1-3(Fucα1-2)Galβ1-3(Fucα1-4)GlcNAc

A6-1
GalNAcα1-3(Fucα1-2)Galβ1-3GlcNAcβ1-3Galβ1-4Glc

Le^b-HSA
Fucα1-2Galβ1-3(Fucα1-4)GlcNAcβ1-3Galβ1-4Glc-conjugated to human serum albumin, average valency: 22

Glc GlcNAc Gal GalNAc Fuc



Structures and synthesis routes for type 1 chain ABO glycans (H and Le^b) that form the BabA receptors on mucosal surfaces of the GI tract. The fucosylation required to obtain H and Le^b antigens occurs on the terminal Gal of the type 1 lacto-series glycan chain. This addition depends on the secretor transferase (FUT2) for the glycans in the saliva, GI mucus secretions and epithelia. A majority of individuals express the secretor enzyme and are fittingly called "secretors". When an α1-4 branched fucose is also added to the core GlcNAc (by FUT3, the Lewis transferase), the Le^b phenotype is expressed. Thus bg O can both constitute the H antigen with only the terminal α1-2 fucose or the Le^b antigen with both the terminal and the α1-4 branched fucoses. These structures form the target glycans for the bg A and B transferases (tA, tB, respectively) that add GalNAc or Gal, respectively, to obtain the A/ALe^b or B/BLe^b bg's. BabA is polymorphic in bg specificity (Aspholm-Hurtig *et al.* 2004), *specialist* BabA bind H and Le^b glycans, *generalist* BabA bind in addition their tA and tB modified derivatives (A, B, ALe^b and Ble^b). In case of non-secretors (ie. individuals lacking FUT2 activity), FUT3 activity on the type 1 core glycan results in Le^a, which is not recognized by BabA. Type 2 chain ABO glycans have a β1-4 Gal-GlcNAc linkage (*) and result in type 2 H and Le^y glycans after FUT2 and FUT3 activity (not shown).

Table S2 **Figures 2, 3, 6, 7**	**BabA^{AD}:Nb-ER14**	**BabA^{AD}:Nb-ER19**	**BabA^{AD}:Nb-ER19 +Le^{b}6**	**BabA^{AD}:Nb-ER19 + ALe^{b}5**	**BabA^{AD}:Nb-ER19 + BLe^{b}7**	**BabA^{AD}:Nb-ER19 + A6-1**	**S831:Nb-ER19**	**S831:Nb-ER19 + Le^{b}6**
Data collection								
Wavelength	0.9791	0.97949	0.96863	0.96863	0.96863	0.96863	0.9786	1.0332
Beamline	PROXIMA-2a	I04	I24	I24	I24	I24	PROXIMA-1	PROXIMA-1
Space group	P 2 2_1 2_1	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1
a, b, c (Å)	42.4, 133.7, 201.6	51.0, 131.7, 123.5	51.1, 131.9, 123.1	50.9, 132.0, 123.8	51.1, 132.4, 123.5	50.9, 131.7, 123.6	51.0, 132.1, 123.5	50.8, 131.9, 125.4
α, β, g (°)	90.0, 90.0, 90.0	90.0, 94.8, 90.0	90.0, 94.8, 90.0	90.0, 94.8, 90.0	90.0, 94.9, 90.0	90.0, 94.7, 90.0	90.0, 94.9, 90.0	90.0, 94.9, 90.0
Resolution range (Å)	47.4-2.7 (2.88-2.74)	58.0-2.2 (2.23-2.17)	45.0-2.7 (2.89-2.72)	33.2-2.3 (2.34-2.28)	43.2-2.8 (2.88-2.81)	34.9-2.4 (2.5-2.44)	48.5-2.6 (2.73-2.59)	48.4-2.4 (2.58-2.44)
Total reflections	180,907 (25,007)	381,170 (29,608)	172,232 (27,643)	410,606 (29,639)	183,559 (13,516)	336,380 (24,962)	179,275 (19,600)	338,859 (44,415)
Unique reflections	31,163 (4,300)	85,215 (6,359)	43,591 (7,029)	72,982 (5,310)	39,799 (2,905)	60,096 (4,387)	49,587 (6,580)	60,611 (8,600)
CC(1/2)	99.5 (91.7)	99.6 (72.3)	98.2 (50.4)	99.7 (66.4)	99.6 (61.4)	99.6 (54.5)	99.6 (53.9)	99.6 (60.3)
R_{merge} (%)	13.9 (55.5)	4.6 (61.3)	16.0 (91.6)	7.9 (74.3)	8.0 (75.1)	9.5 (78.6)	8.5 (79.1)	9.2 (97.0)
$R_{p.i.m.}$ (%)	6.3 (25.2)	3.7 (48.3)	9.5 (55.9)	5.3 (51.5)	6.1 (57.8)	6.8 (55.9)	5.1 (51.7)	4.3 (46.9)
Average I/σI	7.9 (2.7)	15.6 (2.2)	6.7 (2.6)	12.0 (2.2)	12.5 (1.8)	10.6 (2.2)	9.4 (1.3)	10.1 (1.5)
Completeness (%)	99.3 (95.8)	99.4 (100.0)	99.8 (99.8)	98.6 (97.9)	99.7 (99.9)	99.7 (99.7)	97.7 (89.1)	99.5 (97.0)
Multiplicity	5.8 (5.8)	4.5 (4.7)	4.0 (3.9)	5.6 (5.6)	4.6 (4.7)	5.6 (5.7)	3.6 (3.0)	5.6 (5.2)
Wilson B-factor	38.1	40.4	38.1	44.7	59.6	44.7	77.1	70.6
Refinement								
R_{work}/R_{free}(%)	21.22 / 26.49	17.07 / 20.65	17.90 / 21.76	17.08 / 19.71	17.07 / 20.75	17.57 / 20.51	19.47 / 23.03	18.91 / 21.91
average B-factor (Å^2)	27.1	35.8	27.6	36.0	45.5	37.7	46.0	35.7
R.m.s. deviations								
Bond lengths (Å)	0.013	0.019	0.015	0.019	0.014	0.017	0.014	0.016
Bond angles (°)	1,600	1,859	1,646	1,863	1,622	1,741	1,595	1,773
No. Atoms (non H)								
Protein	7445	8031	7974	7976	7970	7979	7972	7986
Carbohydrate	-	-	136	120	158	144	-	136
Water	88	481	150	227	48	137	52	102
Ramachandran								
Favoured (%)	93.77	95.59	94.21	95.61	93.82	95.99	94.21	94.85
Allowed (%)	5.01	3.16	4.55	3.63	4.85	3.44	4.37	3.91
Ouliers (%)	1.23	1.25	1.23	0.76	1.33	0.57	1.42	1.24
MolProbity score	2.26	1.71	1.81	1.73	1.71	1.59	1.85	1.78
PDB entry	**5F7L**	**5F7K**	**5F7M**	**5F7N**	**5F7W**	**5F7Y**	**5F8R**	**5F8R**

Crystallographic data and statistics for structures presented in Figure 2, 3, 6 and 7

	A730:Nb-ER19 + Le^b6	A730:Nb-ER19 + A6-1	P436:Nb-ER19 + Le^b6	P436:Nb-ER19 + BLe^b7
Data collection				
Wavelength	0.98	0.97	0.98	0.93
Beamline	PROXIMA-1	I24	I04	IO4-1
Space group	P 1 2_1 1	P 1 2_1 1	C 2 2 2_1	C 2 2 2_1
a, b, c (Å)	103.2, 134.5, 124.0	102.9, 134.9, 123.8	124.0, 136.2, 124.6	119.2, 135.1, 127.3
α, β, g (°)	90.0, 102.3, 90.0	90.0, 102.5, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	48.64-2.99 (3.16-2.99)	67.46-2.62 (2.76-2.62)	51.54-2.44 (2.57-2.44)	73.15-2.59 (2.66-2.59)
Total reflections	447382 (56631)	558740 (82864)	257262 (36479)	213485 (16651)
Unique reflections	65294 (8335)	98949 (14409)	39582 (5717)	32091 (2365)
CC(1/2)	99.6 (61.5)	99.7 (54.5)	99.6 (70.2)	99.9 (65.7)
R_{merge} (%)	18.0 (135.0)	11.3 (119.7)	9.5 (93.0)	6.7 (116.2)
$R_{p.i.m.}$ (%)	7.4 (55.1)	5.2 (54.3)	4.3 (43.5)	3.1 (50.9)
Average I/σI	9.9 (1.7)	9.6 (1.4)	8.8 (1.7)	15.6 (1.7)
Completeness (%)	97.9 (86.1)	99.8 (99.9)	100.0 (100.0)	99.5 (99.5)
Multiplicity	6.9 (6.8)	5.6 (5.8)	6.5 (6.4)	6.7 (7.0)
Wilson B-factor	52.7	51.5	44.6	57.8
Refinement				
R_{work}/R_{free}(%)	18.11/23.38	19.06/23.69	20.48/25.74	19.50/23.67
average B-factor ($Å^2$)	52.93	44.62	40.8	58.4
R.m.s. deviations				
Bond lengths (Å)	0.012	0.014	0.017	0.016
Bond angles (°)	1,535	1,677	1,877	1,792
No. Atoms (non H)				
Protein	16195	16305	3971	3972
Carbohydrate	272	284	68	79
Water	42	75	45	18
Ramachandran				
Favoured (%)	93.87	93.51	96.39	95.98
Allowed (%)	5.61	5.14	3.23	3.07
Ouliers (%)	0.52	1.35	0.38	0.96
MolProbity score	2.01	1.97	1.57	1.66
PDB entry	**5F93**	**5F97**	**5F9A**	**5F9D**

Table S3- related to Figures 1, 3: MS fingerprint of trypsinized native BabA 17875

Name	Peptide 1	Peptide 2	Mass	M/Z	Charge
CL1	NSPAYQAVLLAINAAVGFWNVLGYAT QCGGNANGQESTSSTTIFNNEPGYR	STSITCSLNR	6414.0	1605.5	4
				1284.8	5
CL2	ENNGTVSVTYTYTCSGEGNDNCSK		2570.0	1271.4	2
CL3	LDGVPDSAQALLAQASTLINTINTACP YFSVTNK	LCQFTEEISAIQK	4971.5	1659.1	3
				1244.9	4
				996.1	5
CL4	NFVTGFLATCNNK	STAGTSGTQGSPPGTVTTQTFASGCAYVEQT ITNLNNSIAHFGTQEQQIQQAENIADTLVNFK	7997.8	1601.6	5
				1334.7	6

MS fingerprints for analysis of disulfide connectivity were performed by ProtTech, Inc.

SUPPLEMENTARY INFORMATION

EXTENDED EXPERIMENTAL PROCEDURES

1. Bacterial Strains and Culture conditions.

1.1. Strains of H. pylori

The *H. pylori* reference strain CCUG 17875 (designated 17875) was used as standard strain and binds both ABO/Le^b and sialylated antigens, such as sLe^x (Mahdavi et al., 2002). The 17875*babA1A2* mutant with deletion mutations of the two *babA* alleles does not bind ABO/Le^b but binds to sialylated antigens (Ilver et al., 1998). The 17875/Leb strain is a spontaneous mutant that binds ABO/Le^b but has lost binding to sialylated antigens (Mahdavi et al., 2002). Strain J166, used for the *CL2* contraselection mutation, is a strain used for Rhesus macaque challenges and is *vacA* (s1), *babA* and cagPAI positive (Solnick et al., 2001). The J166Δ*babA*::*rpsLCAT* (described in section 2.8.) was used as control.

H. pylori S831 is a Spanish specialist strain and clone S831G is a spontaneous generalist derivative identified in this study. Strain P1Δ*babA* was used for shuttle vector complementation (described in section 2.8).

1.2. Culture of H. pylori

H. pylori were grown on BD BBL Brucella agar (Becton, Dickinson and Company, MD, USA), supplemented with 10% of heat-inactivated calf serum (Svenska LabFab, Ljusne, Sweden), 1% Isovitox Enrichment (Dalynn Biologicals, Calgary AB, Canada) and antibiotic mixture (4 μg/μL amphotericin B, 5 μg/μL trimethoprim and 10 μg/μL vancomycin (Sigma Aldrich, St. Louis, MO, USA) under micro-aerophilic conditions (10% CO_2 and 5% O_2) in mixed-gas incubator (Forma Scientific Inc., AB Ninolab, Väsby, Sweden) at 37°C.

2. Nanobody generation, BabA native and recombinant procedures, protein expression and purification.

2.1. Purification of native Leb-binding BabA.

Purification of native BabA from *H. pylori* was essentially as described by Bugaytsova *et al*, (in final preparation); briefly *H. pylori* 17875/Leb bacterial cells were suspended in N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfate (ZW-12; Sigma-Aldrich, St Louis, MO, USA). The protein extract was subjected to ion exchange chromatography and BabA protein was eluted with a linear NaCl gradient, and fractions were assayed for the presence of BabA by immunoblotting with anti-BabA AK277 antibody (Odenbreit et al., 2002) (described in section 3.3). Pooled fractions were applied to a Sepharose 4 Fast flow matrix conjugated with Le^b antigen (Isosep AB,

Tullinge, Sweden). BabA was eluted by a linear acid gradient. BabA purity was analyzed by SDS-PAGE and Coomassie staining.

2.2. Cloning, expression and purification of recombinant full-length BabA.

The coding sequence for the mature BabA coding region (*babA2*) was PCR amplified from *H. pylori* strain CCUG 17875 and cloned into plasmid pASK-HR2, which has been generated by modification of pASK-Iba12 (Iba, GmbH) so that the OmpA leader sequence is replaced with the DsbA leader sequence and with the inclusion of a 6x His tag between the existing *StrepII*-tag and the thrombin cleavage site. For protein expression, an overnight pre-culture of *E. coli* Top10 transformed with pASK-HR2-BabA was inoculated in LB broth (1:100) supplemented with 100 µg/mL ampicillin and grown at 26 °C to an OD_{600} 1 before induction with 0.2-µg/mL anhydrotetracycline. After 6 hours of induction, the cells were recovered, washed in TBS (20 mM Tris pH 8.0, 150 mM NaCl), weighed and stored at -80 °C. For BabA purification, cells were resuspended in TBS supplemented with 5 µM leupeptin and 1 mM AEBSF, 20 µg/mL deoxyribonuclease and 100 µg/mL lysozyme, lysed by a single passage in a Constant System Cell Cracker at 20 kPsi at 4 °C and debris was removed by centrifugation at 10.000*g* for 10 minutes. The supernatant underwent ultracentrifugation at 100.000*g* for 45 minutes. Inner membranes were selectively solubilized by resuspension of total membrane pellets in TBS with 0.5% (w/v) N-laurosarcosine, 5 µM Leupeptin and 1 mM AEBSF. Outer membranes were pelleted from the suspension by ultracentrifugation at 100.000*g* for 45 minutes and subsequently solubilized with 100 mL of 1 % (w/v) N-dodecyl-β-D-maltopyranoside (DDM) in 20 mM Tris pH 8.0, 500 mM NaCl, 5% (v/v) glycerol, 100 µg/mL lysozyme, 5 µM leupeptin and 1 mM AEBSF. Following a final ultracentrifugation step, the outer membrane extract was loaded at 4 °C on a 5 mL Ni-Sepharose 6 Fast Flow resin (GE Healthcare) equilibrated with 5 column volumes of 20 mM Tris pH 8.0, 500 mM NaCl, 5% (v/v) Glycerol, 0.03 % (w/v) DDM supplemented with 20 mM imidazole. The column was washed with 20 column volumes of the equilibration buffer and a linear gradient was applied from 20 to 500 mM imidazole. Elution fractions were further purified by affinity chromatography by using 5 mL Strepactin column (Iba GmbH) equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl, 5% (v/v) Glycerol, 0.03 % (w/v) DDM, 1 mM EDTA and eluted with buffer A supplemented with 2.5 mM of desthiobiotin.

2.3. Generation and selection of anti-BabA llama single domain antibodies or nanobodies

An alpaca was immunized 6 times at weekly intervals with Gerbu LQ 3000 adjuvant supplemented with 150 µg BabA purified from *H. pylori* strain 17875/Leb as a native, detergent-solubilized protein (see experimental procedures 2.1). The anti-BabA nanobody phage-display library was generated according to previously published methods (Ghahroudi et al., 1997). From the lymphocytes extracted from anti-coagulated blood of the immunized alpaca, cDNA was

prepared and used as template in a 2-step nested PCR that amplified the genes coding for the variable domains of the heavy-chain antibodies. The amplified pool of nanobody DNA fragments were ligated into a pHEN4 phagemid vector and transformed into *Escherichia coli* TG1 cells. A nanobody library of 10^9 transformants was generated, with 86% of the clones within the library containing a vector with the nanobody gene insert of the proper size, as determined by PCR. Next, the nanobody repertoire was expressed on M13 virions after infection with M13K07 helper phages and biopanning of phage particles expressing BabA-specific nanobodies was performed. After four consecutive rounds of selection on immobilized antigen, a clear enrichment of phage particles for BabA was observed. The third and fourth round of selection were also carried out at a lower pH (200 mM glycine solution at pH 4.5 and 2.5) than in the standard protocol. After the fourth round of selection, 428 colonies from the three different selection conditions were randomly chosen for expression of their nanobody as soluble protein. When the crude nanobody periplasmic extracts were tested in an ELISA, 345 out of the 428 extracts were shown to be specific towards BabA. Based on optical density measurements of the ELISA, 24 colonies from the second and third rounds of all three selection conditions were sent for DNA sequencing. 21 unique anti-BabA nanobody DNA fragments were re-cloned in the expression vector pHEN6 to contain a His_6 tag facilitating the purification of the recombinant nanobody and transformed into *E. coli* WK6 cells, where nanobodies are transported into the periplasm, upon expression. Nb-ER14 and Nb-ER19 were selected for use in structural studies on the basis of their stabilizing properties towards the $BabA^{AD}$ fragment (see extended experimental procedures 4.2).

2.4. Expression and isolation of nanobodies Nb-ER14 and Nb-ER19

Overnight precultures of *E. coli* WK6 Su- cells harboring the pHEN6 vector with insert encoding for either 6xHis-tagged Nb-ER14 or Nb-ER19 was used to inoculate lysogeny broth (LB) media (1:100) supplemented with 100 μg/mL ampicillin. Bacterial cells were grown at 37°C, induced at OD_{600} of ~ 0.8 with 1.0 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cell cultures incubated overnight at 20°C while shaking. The cell pellets were harvested by centrifugation at 6238 rcf for 15 minutes at 4°C. Extraction of periplasmic proteins was performed by suspending every 1 g of wet cell pellet in 4 mL of 20 mM Tris-HCl pH 8.0, 20 % (w/v) sucrose buffer supplemented with protease inhibitors AEBSF-HCl (0.4 mM) and leupeptine (1 μg/mL). The mixture was left on ice for 20 minutes together with 5 mM EDTA and 17,560 U of lysozyme. $MgCl_2$ (40 mM final concentration) was then added and the periplasmic extract was collected by centrifugation at 17,418*g* for 20 minutes at 4°C.

2.5. $BabA^{AD}$ expression, Nb complex formation and purification

The cloning and expression of $BabA^{AD}$ as a stable, soluble BabA fragment are described in Subedi *et al.* (Subedi et al., 2014). In the resulting expression plasmid (pASK-HR3-$BabA^{25-460}$), the coding sequence for residues 25 to 460 of mature BabA (excluding the leader sequence) from *H. pylori* CCUG 17875 and a C-terminal 6xHis tag are cloned in a pASK-IBA12 derivative vector (pASK-HR3) in which the *ompA* leader sequence was replaced by the *dsbA* leader. For $BabA^{AD}$ expression, an overnight preculture of *E. coli* Top10 cells containing pASK-HR3-$BabA^{AD}$ was used to inoculate LB medium (1:100) supplemented with 100 μg/mL^{-1} ampicillin. Bacterial cells were grown at 37°C while shaking until the cultures reached an OD_{600} of ~ 0.5 upon which temperature was decreased to 20°C. At an OD_{600} of 1.0 the bacterial cell cultures were induced with anhydrotetracycline (0.2 mg/L) for overnight expression at 20°C. Cells were harvested by centrifugation at 6238*g* for 15 minutes at 4°C and periplasmic extraction was performed as described above for Nb-ER19 or Nb-ER14.

The periplasmic extracts of $BabA_{1-460}$ and either Nb-ER14 or Nb-ER19 were mixed in a ratio of 5:1 (v/v) on ice, filtrated through a 0.45 μm filter and sodium chloride was added to a final concentration of 0.5 M prior to loading on a 5 mL pre-packed Ni-NTA column (GE Healthcare) equilibrated with buffer A (20 mM Tris-HCl pH 8.0, 0.5 M NaCl and 5 % (v/v) glycerol). The column was then washed with twenty-five bed volumes of buffer A and bound proteins were eluted with a linear gradient of 0-60 % buffer A supplemented with 500 mM imidazole. Fractions containing the protein complex, as determined in SDS-PAGE, were pooled and concentrated in a 10 kDa MW cutoff spin concentrator to a final volume of 500 μL. To remove excess unbound Nb and minor protein contaminants, the concentrated sample was injected onto the Hi-Prep™ 26/60 Sephacryl S-100 HR column (GE Healthcare) pre-equilibrated with a buffer containing 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 % (v/v) glycerol. Fractions containing the $BabA^{AD}$:Nb-ER19 complex were pooled and concentrated to a final concentration of 15-20 mg mL^{-1} using a 10 kDa MW cutoff spin concentrator.

2.6. Production of selenomethionine labeled $BabA^{AD}$:Nb-ER14

To produce selenomethionine labeled $BabA^{AD}$ the methionine auxotrophic *E. coli* strain B843 was transformed with pASK-IbaH3-$BabA^{AD}$ and precultured in M9-based minimal media, comprised of SelenoMet Nutrient Mix (1:200; w/v) (Molecular Dimensions), SelenoMet Medium Base (1:50; w/v)(Molecular Dimensions), L-methionine (40 μg/mL) and ampicillin (100 μg/mL). The overnight culture was pelleted, resuspended in an equal volume 1x PBS and used to inoculate the above minimal medium (1:100 dilution) supplemented with L-selenomethionine (40 μg/mL culture)(Acros Organics) instead of L-methionine. Protein purification and Nb complex formation were performed as described for unlabeled protein (see above), but with the addition of 5 mM β-mercaptoethanol in the buffers used to prevent selenomethionine oxidation.

Prior to crystallization, the selenomethionine labeled $BabA^{AD}$ in complex with unlabeled Nb-ER19 was concentrated to a final concentration of 20 mg/mL using a 10 kDa MW cutoff spin concentrator.

2.7. Cloning and production of the $BabA^{AD}$ S831 DL1 graft mutant

For the generation of $BabA^{AD;DL1\text{-}S831}$ the DL1 loop region of the generalist *H. pylori* strain CCUG 17875 (residues 198-208: SKKATGVSDQN) were substituted with the DL1 sequence of specialist *H. pylori* strain S831 (residues: LPKVTGVDNQN). Substitution of the DL1 coding sequences was performed using site directed, ligation independent mutagenesis (SLIM) as described by Chiu et al. (Chiu et al., 2004). Primers used are F_T: TTGCCAAAAGTCACAGGTGTAGACAACCAAAACGGCGGAACCAA AAC, F_S: GGCGGAACCAAAAC, R_T: GTTTTGGTTGTCTACACCTGTGACTTTTGGCAAGCAGTTATCATTCCC, and R_S: GCAGTTATCATTCCC. The PCR reaction mixture was subjected to DpnI cleavage to selectively digest the CpG methylated parent pASK-HR3-$BabA^{25\text{-}460}$ and was transformed into *E. coli* Top10. $BabA^{AD;DL1\text{-}S831}$ and its complex with Nb-ER19 were produced as described above for $BabA^{AD}$.

2.8. Cloning and production of the $BabA^{AD}$ A730 and P436 CBD exchange mutants

In order to generate $BabA^{AD;\ ID\text{-}730}$ and $BabA^{AD;\ ID\text{-}P436}$ the carbohydrate binding domain of the generalist *H. pylori* strain CCUG 17875 (residues 180-257) was exchanged for the CBD's of the Alaskan A/B specialist strain A730 and the Peruvian generalist strain P436. The creation of such hybrid BabA variants will have the advantage that any phenotypic changes are a function of the CBD only, and also this approach allows nanobody aided crystallization of the hybrids since the above described nanobodies have been shown to recognize the conserved BabA core domain. First the 17875 CBD was removed and a BamHI restriction site inserted between residues 179 and 258 by performing a standard PCR with primers F: GGCATCGGATCCGTTGTTTTCTTTGAGCG CGGGTAAGCC and R: ATCGCAGGATCCGCTCAAGCGCTCTTGGCGCAAGCG on template plasmid pASK-HR3-$BabA^{25\text{-}460}$. The PCR reaction mixture was subjected to DpnI cleavage to selectively digest the CpG methylated parent plasmid and was transformed into *E. coli* Top10. Afterwards the resulting plasmid was cleaved with subsequently BamHI and RecJf restriction enzymes to create a linear blunt ended DNA fragment. Synthetic DNA corresponding to the CBD's of strains A730 and P436 (Gen9) was recombined into the linear DNA fragment with In-Fusion (Clontech). $BabA^{AD;\ ID\text{-}730}$ and $BabA^{AD;\ ID\text{-}P436}$ and their complexes with Nb-ER19 were produced as described above for $BabA^{AD}$.

2.9. Creation of CL2 mutants by contraselection and shuttle vector mutagenesis

To generate an isogenic mutant of strain J166 in which the BabA CL2 loop cysteines were mutated to Ala (eg. Cys189A/Cys197A), we used a contraselection procedure as was described (Dailidiene et al., 2006). Briefly *H. pylori* strain J166 was made streptomycin resistant by transformation of chromosomal DNA from a 26695 *rpsL* mutant strain previously described (Barrozo et al., 2013), obtaining J166Str^r. Next, part of the *babA* promotor and the part of the ORF downstream of the CL2 region in the J166Str^r strain were replaced with a two-gene *rpsL*CAT cassette plasmid, which encodes dominant streptomycin susceptibility (*rpsL*) and chloramphenicol resistance (*CAT*). The *rpsLCAT* cassette was amplified with forward primer rpsL_F_SacI (5'-AACGAGCTCGATGCTTTATAACTATGGATTAAACAC-3') and reverse primer C2-CatR_BamHI (5'-AACGGATCCTTATCAGTGCGACAAACTGGGAT-3'), and then ligated to J166 *babA* DNA for the "Left arm" with forward primer HP0898F-NotI (5'-AATGCGGCCGCCAAAGAAGCCAAGCTAGAAATCC-3') and reverse primer BabAupR_SacI (5'-AACGAGCTCTTTGTTTCGTTTGGTGGTAAA-3'), and "Right arm" with forward primer BabAF4_BamHI (5'-AACGGATCCGCGAGCACGCTCATTAACACC-3') and reverse primer HP1243:1862L26_Xho (5'-AACCTCGAGTTGGTAGCTTTATCGTTGATGAAGTT-3'). The three pieces where cloned into a pBluescript vector and used to transform the J166Str^r strain, making the strain sensitive to streptomycin but resistant to chloramphenicol. Thereafter a DNA fragment containing the *babA*-CL2 mutant gene was obtained by creating a "Left arm" with forward primer HP0898:1029U23 (5'-TTGCACCCCACAAAACCCGATTG-3') and reverse primer J166_BabA_AlaR (5'-GGCGTTATTATTCCCTTTTCCTGAGGCTGTGTA-3'), a "Right arm" with forward primer J166_BabA_AlaF (5'-AGCCTCAGGAAAAGGGAATAATAACGCCTC-3') and reverse primer HP1243:1895L30 (5'-CCCACAGAAAGCTTGTTGTTTTTGCCTAAG-3') and subsequently joining these two fragments in a PCR reaction with forward primer HP0898:1037U20 (5'-CACAAAACCCGATTGGCAGT-3') and reverse primer HP1243:1862L26_Xho (5'-AACCTCGAGTTGGTAGCTTTATCGTTGATGAAGTT-3'). This DNA fragment was used to transform the J166*rpsLCAT* strain, replacing the *rpsLCAT* cassette, making the strain streptomycin resistant once more, and thus obtaining the J166*CL2* mutant.

Entire *babA* sequences in potential *CL2* mutants were obtained by sequencing at MWG Operon, Germany using primers: J166babA_Seq4 (5'-CTTCTGACGTATGGACTTATGG-3'), J166_CysF (5'-AGCCTCAGGAAAAGGGAATAATAACGCCTCGGAAC-3'), J166_CysR (5'-GTTCCGAGGCGTTATTATTCCCTTTTCCTGAGGCT-3'), NdeI_O2babAF (5'-GATCCATATGAAAAAAACTCTTTTACTCTCTC-3'), NdeI_O2babAR (5'-CAAGGCGGCCGCTTAATAAGCGAACACATAGTTCAAATAC-3'). The contraselection approach in J166 generally produced frequent random point mutations. Clone 33 referred to as J166*CL2* held the correct gene sequence in the adhesion domain, with three non-silent point mutations outside our region of interest: Glu399Gly, ILe440Phe, Ser660Asn. As control, additional CL2 mutants in

J166 were made were either Cys residue was replaced with Ala and the entire BabA sequence was without additional mutations, which confirmed the non-binding phenotype. These strains were constructed with a chloramphenicol marker behind the *babA* gene, which proved less mutagenic, but not completely isogenic to the WT as the J166*CL2* Clone 33, but verified the CL2 importance (data not shown). *H. pylori* strain P1Δ*babA* was genetically complemented with the shuttle vector (SV) pIB6 (Jiménez-Soto et al., 2013) containing insert *babA* and *babA*-*CL2* genes from strain 17875, essentially as described in Bugaytsova *et al*. (in preparation). GenScript USA Inc. performed *babA* DNA synthesis, Cys/Ala-mutagenesis and subcloning into the SV. Plasmid variants were introduced into the P1Δ*babA* strain via bacterial conjugation by using *E. coli* counterselection with diaminopimelic acid and kanamycin selection for the SV. The SV plasmid was re-isolated after introduction into P1Δ*babA* by using E.Z.N.A. Plasmid Mini Kit I (OMEGA Bio-tek) and sequencing to confirm correct insert was made at MWG Operon, Germany. Shuttle vector over-expression clones inherently vary in BabA expression and generally BabA expressing clones loose Leb-binding over time (our unpublished finding), e.g. P1Δ*babA* expressing the *CL2* 17875 *babA* showed between 27 % and 0.5 % binding (data not shown), an experiment of 9 % binding is shown in Figure 4B, where BabA levels of WT and *CL2* 17875 *babA* expressed at similar levels as judged by an immunoblot. The system is a compromise where qualitative binding properties of BabA-*CL2* (as probed with SPR) can be performed (Figure 4C), but detailed quantification is not meaningful due to individual and time-dependent variation of the plasmid-based expression.

3. Protein and sample characterization

3.1. Oligomerization of the BabA protein detected by in vivo cross-linking on H. pylori cells.

For cross linking of BabA oligomers on the bacterial cell surfaces, *H. pylori* 17875/Leb cells were harvested after 24 h of incubation and first washed twice in PBS-Tween. Then, 1 mL of bacterial suspension (OD_{600} = 1.0) was pelleted for 5 min at 5000*g* and re-suspended in 40 µL of citrate-phosphate buffer, pH 5.5. After 10 min of incubation at T_r (room temperature), 10 µL of the cell suspension was mixed with 10 µL of the citrate-phosphate buffer, pH 5.5 and 5 µL of 0.25% glutaraldehyde (Scharlau Chemie, EU). Sample was incubated at T_r for 10 min and then mixed with equal volume of 2 x Laemmli buffer under non-reducing conditions. Oligomerization was visualized by immunoblotting as described in 3.3.

3.2. Thermal denaturation assays

Solutions of 7.5 µL of 300 x SYPRO Orange (Invitrogen), 5 µL of 2.5 mg/mL $BabA^{AD}$ or $BabA^{AD}$:Nb-ER19 and 12.5 µL of buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 % (v/v)

glycerol) were added to the wells of a MicroAmp Optical 96-well Reaction Plate (Applied Biosystems). The plates were sealed with Microamp optical adhesive film (Applied Biosystems) and heated in a 7500 Real Time PCR System (Applied Biosystems) from 20 to 95 °C in increments of 1 °C. Fluorescence changes in the wells of the plate were monitored simultaneously and plotted against temperature. The wavelengths for maximal excitation and emission of the SYPRO orange are 470 and 570 nm, respectively.

3.3. Immunoblot

Immunoblotting was performed essentially as described previously (Bäckström et al., 2004). Protein extracts were separated on SDS-PAGE using 7.5% Tris-HCl Ready Gels (Bio-Rad Laboratories, Hercules, CA, USA) or 4-12% Bis-Tris Ready Gels (Thermo Fisher Scientific, Life Technologies) followed by polyvinylidene difluoride (PVDF) membrane immunoblotting (Bio-Rad, USA). The membrane was blocked in 5% defatted milk in PBS-Tween followed by anti-BabA AK 277 antibody (1:6000) (Odenbreit et al., 2002) incubation and then probing with anti-rabbit-Ab-HRP (1:1000) (DakoCytomation, Denmark A/S). Signals were developed with ECL chemiluminescence (SuperSignal West Pico, Thermo Scientific/Pierce, IL, Rockford, USA) on High Performance Chemiluminescence film (GE Healthcare) Precision Plus Protein™ Prestained Standards (Bio-Rad Laboratories, Hercules, CA, USA) were used. In Figure S4F, *H. pylori* from strains J166 and J166*CL2* were diluted in sample buffer for equal concentration after OD_{600} measurement and extracts were either incubated for 30 min at 37°C (- -), treated with 5 % β-mercaptoethanol (+ -), 5 min boiling (- +) or both the later treatments (+ +). Samples were next subjected to SDS-PAGE and immunoblotting with AK277.

3.4. Mass spectrometry of Cys-loop connectivity in native BabA

The protein identification and disulfide mapping of native BabA was carried out at ProtTech, Inc. by using the NanoLC-ESI-MS/MS peptide sequencing technology. In brief, native BabA purified from strain 17875/Leb was digested in-gel with sequencing grade modified trypsin. The resulted peptides mixture was analyzed by a LC-ESI-MS/MS system, in which a high performance liquid chromatography (HPLC) with a 75 micrometer inner diameter reverse phase C18 column was on-line coupled with an ion trap mass spectrometer. The mass spectrometric data acquired were used to map the protein sequence with ProtTech's proprietary software suite (www.ProtTech.com). For mapping dissulfide connectivity, native babA was incubated with iodoacedimide and then dialyzed with MWCO 8,000 membrane overnight before digestion with sequencing grade modified trypsin. The resulted peptides mixture was analyzed by a LC-ESI-MS/MS system as above. The mass spectrometric data acquired were analyzed with ProtTech's proprietary S2MAP (version 1.3) software suite.

3.5. Mass spectrometry of N-acetylcysteine titration of BabA^{AD}

An N-acetyl-L-cysteine (NAC; Sigma-Aldrich) stock solution was prepared at 100 mg/mL in 50 mM sodium phosphate pH 7.0, 150 mM NaCl and the pH adjusted to 7.0. A two-fold dilution series of NAC was prepared and mixed in 1:1 volume ratio with 0.8 mg/mL BabA^{AD} (50 mM phosphate pH 7.0, 150 mM NaCl). The protein solution was digested overnight at 37 °C by trypsin (sequencing grade modified trypsin, Promega) at a 1:25 trypsin/protein mass ratio. Before analysis by mass spectrometry, the samples were desalted on ZipTip C18 (Millipore) and eluted in 50% acetonitrile /1% formic acid (v/v). The samples were loaded into a nanoflow capillary (Thermo Scientific) and ESI mass spectra were acquired on a quadrupole time-of-flight instrument (Q-Tof Ultima, Waters) operating in the positive ion mode, equipped with a Z-spray nanoelectrospray source. Data acquisition was performed using a MassLynx 4.1 system. The reduction of the Cys189/Cys197 disulfide containing peptide Glu176-Lys199 was monitored by mass spectrometry by following the evolution of the intensity of the peaks at m/z 1271.03 and 1271.53, belonging to the isotopic distribution of the doubly-charged peptide in the oxidized form. The intensities of both peaks were normalized with respect to the most intense peak of the isotopic distribution. These peaks were chosen because of their measurable intensity in all the spectra and because these do not interfere with the isotopic distribution of the reduced form (see Figure S5B and S5C).

3.6. Trypsination of BabA in H. pylori J166 and J166CL2

An overnight suspension of *H. pylori* strain J166 was adjusted to OD_{600} of 0.5 and 200 µL was treated with 50 µg/mL trypsin (Invitrogen) for 2 and 30 minutes. The trypsination mix was then quenched by 100 µL fetal calf serum and immediately placed on ice. Bacterial cells were pelleted, resuspended in sample buffer and extracts were separated by SDS-PAGE and visualized by immunoblotting with antibody α-BabA AK277 (Figure S4G).

3.7. Circular Dichroism (CD) measurements

BabA^{AD} samples in absence and presence of 100 mM tris(2-carboxyethyl)phosphine (TCEP) were prepared at a concentration of 3 mg/mL in a buffer containing 20 mM Tris pH 8.0, 150 mM NaCl and 5% glycerol. The samples were placed in a 0.1-mm path-length quartz cuvette in a J-715 spectropolarimeter (JASCO). For each sample, three scans were recorded between 250 and 190 nm at a scan speed of 20 nm min^{-1} and averaged to obtain the CD profile. After CD analysis these samples were analyzed with SDS-PAGE in absence of any reducing agents in the loading dye.

4. Structure determination

4.1. Crystallization and structure determination of BabA^{AD} complexes with Nb-ER14 and Nb-ER19

Native and selenomethionine labeled BabA^{AD}:Nb-ER14 complexes were concentrated to 20 mg/mL and crystallized by sitting drop vapor diffusion at 20 °C using 0.2 M Lithium sulfate, 0.1 M Sodium acetate 4.5, 50 % w/v PEG 400 as crystallization buffer. Data were collected at -196°C using loop-mounted BabA^{AD}:Nb-ER14 crystals that were first flash-cooled in liquid nitrogen. Data for selenomethionine- and unlabeled BabA^{AD}:Nb-ER14 crystals were collected at the Proxima-1 beamline (SOLEIL, Gif-sur-Yvette, France) at a 0.98 Å wavelength and integrated and scaled to, respectively, 3.17 and 2.75 Å resolution using the XDS package (Kabsch, 1993). Crystals are in space group $P22_12_1$ with unit cell dimensions a= 42.4 Å, b= 133.7 Å and c= 201.6 Å, corresponding to two BabA^{AD} and two Nb-ER14 copies per asymmetric unit (AU). The unmerged data from the selenomethionine-labeled BabA^{AD}:Nb-ER14 crystal were used for phase calculation using a single anomalous dispersion (SAD) methodology. For this, data normalization, location of 10 selenium atoms in the AU and phase calculation and density modification were performed using the programs SHELXC, SHELXD and SHELXE, respectively (Sheldrick, 2010). The structure factors for the unlabeled BabA^{AD}:Nb-ER14 crystal were phased by molecular substitution with experimental and solvent flattened SAD phases and extended to 2.75 Å using the program DM (Cowtan, 1994; Winn et al., 2011). An initial model was build using the program ARP/wARP (Langer et al., 2008), followed by iterative cycles of manual rebuilding in the graphics program COOT (Emsley and Cowtan, 2004) and maximum likelihood refinement using Refmac5 (Murshudov et al., 2011; Winn et al., 2011), resulting in a final model holding two copies of the BabA^{AD}:Nb-ER14 complex (comprising residues 57-459 and 1-121 for BabA^{AD} and Nb-ER14, resp.) with an R and freeR factor of 21.49 and 26.49 %, resp. (See Table S1 for full processing and refinement statistics).

4.2 Crystallization and structure determination of BabA^{AD}:Nb-ER19, BabA$^{AD;\ ID\text{-}730}$:Nb-ER19, BabA$^{AD;\ ID\text{-}P436}$:Nb-Er19, BabA$^{AD;S831\text{-}DL1}$:Nb-ER19 and their complexes with blood group antigens

BabA^{AD}:Nb-ER19, BabA$^{AD;\ ID\text{-}730}$:Nb-ER19, BabA$^{AD;\ ID\text{-}P436}$:Nb-Er19 and BabA$^{AD;S831\text{-}DL1}$:Nb-ER19 were concentrated to 20 mg/mL and crystallized by sitting drop vapor diffusion at 20°C using 0.2M sodium nitrate, 0.1M Bis Tris propane pH 6.5 and 20% w/v PEG 3350 (for BabA^{AD}:Nb-ER19 and its complexes with bg ligands); 0.1 M isopropanol, 0.1 M HEPES pH 7.5, 20% ethylene glycol and 10 % w/v PEG8000 (for BabA$^{AD;\ ID\text{-}A730}$:Nb-ER19); 1 M Sodium acetate, 0.1 M Imidazole pH 8.0 (for BabA$^{AD;\ ID\text{-}P436}$:Nb-ER19 in complex with HLeb6)or 0.2 M potassium thiocyanate, 0.1 M Bis Tris propane pH 7.5 and 20 % w/v PEG 3350 (for BabA$^{AD;S831\text{-}DL1}$:Nb-ER19 and its Le^b6 complex) as a crystallization buffer. Complexes of BabA^{AD}:Nb-ER19 and BabA$^{AD;S831\text{-}DL1}$:Nb-ER19

with the different blood group antigens (see table S1) were obtained by addition of 1 mM of the respective bg glycans prior to crystallization. Glycans used for co-crystallization are: bg A Le^b pentasaccharide (ALe^b5, Table S1) and bg A type 1 hexasaccharide (A6-1, Table S1), H Le^b hexasaccharide (Leb6, Table S1), bg B Le^b heptasaccharide (BLeb7, Table S1). Crystals were soaked in crystallization buffer supplemented with 15 % (v/v) glycerol, loop-mounted and flash-cooled in liquid nitrogen. Data were collected at 100 K at beamlines I04, I24 (Diamond Light Source, Didcot, UK) and Proxima 1 (SOLEIL, Gif-sur-Yvette, France; see Table S1 for details) and were indexed, processed and scaled using the XDS package (Kabsch, 1993). All crystals were in the $P2_1$ space group with approximate unit cell dimensions of a=51.0 Å, b=131.7 Å, c=123.5 Å and beta=95° (see Table S2 for details) and two copies of the $BabA^{AD}$:Nb-ER19 complexes per AU. Phases were obtained by molecular replacement using the $BabA^{AD}$:Nb-ER14 structure and the program phaser (McCoy et al., 2007; Winn et al., 2011). The models were refined by iterative cycles of manual rebuilding in the graphics program COOT (Cowtan, 1994; Winn et al., 2011) and maximum likelihood refinement using Refmac5 (Murshudov et al., 2011; Winn et al., 2011). Glycan ligands were refined using the standard geometry restraints incorporated in Refmac5. Table S2 summarizes the crystal parameters and data processing statistics for all the structures.

5. Binding assays

5.1 Bacterial cells and gastric mucosa tissues used for fluorescent microscopy tests.

H. pylori cells used in fluorescent microscopy studies were harvested after 24 h of incubation, washed three times in PBS-Tween and if needed were labeled by FITC (Sigma, St. Louis, MO) as described previously (Aspholm et al., 2006). Then, bacteria were re-suspended in 1 mL of SIA blocking buffer (1% of periodate oxidized BSA in PBS-Tween-20, pH 7.4) to OD_{600} = 0.2. All gastric histo-tissue sections used in this study were obtained from a healthy patient donor with no gastric disease. Dako Fluorescent Mounting Medium (Dako North America., Inc. CA, USA) was applied to slides prior to the microscopy. Binding was digitalized by a Zeiss AXIOcam MRm microscope (Carl Zeiss AB, Stockholm, Sweden) with a 200x optical magnification and quantified with by Zeiss ZEN 2011 (*blue edition*) software.

5.2. Binding of nanobodies Nb-ER14 and Nb-ER19 to H. pylori cells

Nb-ER14 and Nb-ER19 were labeled with fluorescein isothiocyanate (Invitrogen) according to manufacturer's protocol. Excess of unreacted FITC dye was removed by a NAP-5 Sephadex G-25 column (GE Healthcare).

Suspensions of *H. pylori* strains 17875/Leb and 17875*babA1A2* were pelleted by 5 min, 5000g and resuspended in 500 µl of SIA buffer. Cells were then mixed with 1µg/mL of either of the Nb-ER14 or Nb-ER19 FITC labeled nanobodies. Incubations were performed for 2 hours at T_r in the

dark. The suspensions were centrifuged 5min, 5000*g* and the pellets were rinsed with PBS-0.05% Tween and then resuspended in 50µl of SIA buffer. Dako Fluorescent Mounting Medium was applied to 20 µl of the bacterial suspension and bound FITC-Nb-ER14 and FITC- Nb-ER19 were visualized by fluorescence microscopy (as described above in 5.1)

5.3. In vitro inhibition and detachment of *H. pylori* adherence to human gastric mucosa tissue by NAC.

Suspensions of FITC labeled *H. pylori strain* 17875/Leb were mixed with a series of N-acetylcysteine (Acetylcystein, Meda, 200 mg/mL, solution for nebulizator, pH 7) dilutions prepared in SIA blocking buffer (0, 10, 20 and 100 mg/mL) or with 10 µg/mL of Le^b-HSA conjugate (as positive control for inhibition of binding). Bacteria were incubated at 37 °C for 1 h and then applied to human gastric mucosa histo tissue sections for 2 hour. Unspecific binding was removed by PBS-Tween and slides were subjected to microscopy for digital quantification. The level of inhibition derived from number of remaining attached bacterial cells was visualized and quantified.

In the "detachment" regime (Figure 5A iv, v and vi), FITC labeled *H. pylori* strain 17875/Leb cells were applied to histo tissue sections at room temperature. After 2 h, nonadherent bacteria were removed by PBS-Tween and slides were incubated for 1 hour at 37 °C with a series of NAC dilutions prepared in SIA blocking buffer (10, 20, 100 and 200 mg/mL) or 10 µg/mL of Leb-conjugate. Slides were again washed in PBS-Tween and analyzed as described above in 5.1.

To asses the effect of NAC treatment on receptor availability in the human gastric mucosa histo tissue sections, sections were pre-incubated with a series of N-Acetylcysteine (Acetylcystein, Meda, 200 mg/mL, solution for nebulizator, pH 7) dilutions prepared in SIA blocking buffer (0, 10, 20, 100 and 200 mg/mL NAC) for 2 h at room temperature and then washed 3 times in PBS-Tween. Suspensions of FITC labeled *H. pylori strain* 17875/Leb were applied to NAC-(pre)treated sections and incubated for 1 hour (as above). Nonspecific binding was removed by PBS-Tween and slides were subjected to microscopy as described in 5.1.

5.4. Identification of H. pylori S831 generalist variants in the original stock smear of specialist strain S831.

Suspensions of *H. pylori strain* 17875/Leb or strain S831 were centrifuged for 5 min at 5000*g*. Pellets were then resuspended in 200 µL of SIA blocking buffer and mixed with 2 µg of Alexa488 (green) labeled Le^b-HSA or Alexa555 (red)-labeled ALe^b-HSA conjugates. Fluorochromes Alexa Fluor 488 (Alexa488) and Alexa Fluor 455 (Alexa555) were both from Molecular probes, Life Technologies Corp., Oregon, USA). Leb- and ALeb-conjugates were fluoro-labeled as described

previously (Aspholm et al., 2006). After 2 hours of incubation at room temperature in darkness, the suspensions were centrifuged for 5 min at 5000*g*. The pellets were resuspended in 50 µL of SIA blocking buffer, 15 µL of the suspensions were then applied to microscope slide. Bacterial cells with bound fluorescent glyco-conjugate were counted manually whereas the numbers of non-labeled (non-binding) bacterial cells were estimated by phase contrast microscopy.

To determine the presence of residual S831 generalist derivatives, a colony screening was performed essentially as previously described by (Aspholm et al., 2006), with small modifications. Briefly, S831 wild-type *H. pylori* bacteria, diluted in Brucella broth, were cultured on tetrazolium-containing blood agar plates at 500-1.000 colonies per plate. Colonies were transferred onto nitrocellulose paper, washed 3 times with PBS-tween, and blocked over-night (ON) with 1.2% gelatin/0.5 % BSA in TBST (Tris-buffered saline and Tween 20). The next day, the blocking buffer was removed and membranes were incubated with biotinylated-ALe^b (2µg/membrane) in blocking buffer for 2h at RT, washed 3 times with TBST and then incubated with streptavidin-POD (1:5,000; POD: horse radish peroxidase) for 1h at RT. Then the membrane was washed 3 times with TBST and developed with 4-chloro-1-Naphtol tablets (Sigma-Aldrich). The appearance of a dark spot on the membrane indicates a positive reaction, whereas if a spot appears red it indicates no binding (a negative Tetrazolium reaction). *H. pylori* strain 17875 (Ilver et al., 1998) and the 17875*babA1A2* mutant were used as positive and negative ALe^b binding controls, respectively. The screening revealed 6 dark clones, i.e. generalist derivatives denoted S831G clones 1-6, and 3 red clones denoted S831S clones 1-3. RIA established binding properties as follows. S831G clones 1-6 bound Le^b (%): 40.6, 39.3, 39.2, 35.6, 39.4, 38.8 (Mean=38.8) and bound ALe^b (%): 16.8, 16.2, 15.6, 16.0, 18.2, 15.2 (Mean=16.3). S831S clones 1-3 bound Le^b (%): 30.7, 22.4, 25.1 (Mean=26.1) and ALe^b (%): 0.3, 0.3, 0.2 (Mean=0.3). Thus all S831G clones and S831S clones showed a most homogenous binding pattern. For comparison control strain 17875/Leb bound 33.1 % Le^b and 27.7 % ALeb. Clones number 1 are shown in Figure 7D and F. Binding values for S831 were taken from Aspholm *et al*., 2004. The *babA* genes from S831G clones 1 and 2 with generalist binding phenotype and S831S clones 1 and 2 with specialist binding phenotype were amplified and sequenced using *babA* gene loci specific primers according to (Colbeck et al., 2006).

5.5. Surface plasmon resonance (SPR) measurements

SPR experiments on BabA^{AD}, BabA$^{AD;S831\text{-}DL1}$ and full length recombinant BabA were carried out using a Biacore T200 instrument (GE healthcare). The surface of a CM5 sensor chip (GE healthcare) was activated with a 1:1 mixture of 0.1 M N-hydroxysuccinimide (NHS) and 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC). After activation of the

surface the glycoconjugate of human serum albumin and Le^b pentasaccharide in 10 mM sodium acetate, pH 4 was injected to flow cell 2 (FC2) to be immobilized on the sensor surface via remaining primary amine groups present on HSA. As a control the same amount of non-conjugated HSA was immobilized on FC1. Residual unreacted active ester groups were blocked with 1 M ethanolamine-HCl, pH 8.5. $BabA^{AD}$ was flown over the chip surface with concentrations ranging from 500 nM to 1 nM in HBS buffer (10 mM HEPES, 150 mM NaCl, 1 mM EDTA, 5 % (v/v) glycerol, 0.005 % Tween20, pH 7.4) at a flow rate of 20 µL/min at 25°C. To obtain binding dissociation profiles of the different $BabA^{AD}$ oligomeric states, the protein was run over a HiPrep 16/60 Sephacryl S100 (GE healthcare) size exclusion column equilibrated in HBS buffer and fresh fractions of both the oligomer and monomer peaks ((o) and (m), respectively (see Figure 1 and S1) were injected over the HSA-Le^b5 coated surface.

For binding assessment of full-length recombinant BabA, DDM-solubilized BabA (residues 1-721) extracted from the *E. coli* outer membrane and purified as described above was dialyzed against a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 0.005 % (v/v) Tween-20, 0.05 % (w/v) DDM, 2 % (v/v) Glycerol. BabA concentration series ranging from 500 to 1 nM were injected over the surface for 30 min at a flow rate of 2 µL/min. Association and dissociation constants were fitted with the BIAevaluation software and using a 1:1 Langmuir binding model.

For the inhibition experiments of the S831 DL1 graft mutant $BabA^{AD;S831\text{-}DL1}$, the protein was pre-incubating with Le^b5 or ALe^b5 (Elicityl, France) in a stepwise two-fold dilution series from 5 mM to 1 µM. After a 5 minute incubation, the respective mixtures were injected over the HSA-Le^b coated CM5 chip for analysis of residual binding of $BabA^{AD;S831\text{-}DL1}$.

Experiments with *H. pylori* P1*ΔbabA::17875babA* and its isogenic *CL2* mutant (Cys189Ala/Cys197Ala) were performed on the ProteOn XPR36 (Bio-Rad, CA, USA) SPR-system. Here we used GLM chips (Bio-Rad) to immobilize Le^b-HSA at pH 4.0 as described by the manufacturer. Bacterial cells were kept in PBS with 0.005 % Tween-20 and injected at 25°C using OD 0.2, 0.1 and 0.05 for 10 mins at flow rate 25 µl/min and 10 mins dissociation. The reference cell was coated with HSA-Lea (HSA-conjugated Lewis A), and was subtracted from the analysis flow cell.

5.6. Isothermal titration calorimetry

ITC measurements were performed on a MicroCal iTC200 calorimeter (Malvern). In each case, $BabA^{AD}$ was used as a complex with the stabilizing Nb-ER19 to avoid protein aggregation due to stirring in the calorimeter cell. $BabA^{AD}$:Nb-ER19 (70 µM) was loaded into the cell of the calorimeter and the Le^b blood group antigens (eg. Le^b5 or BLeb7, see above) were loaded in the syringe at 1.5 mM concentration. All measurements were done at 25°C, with a stirring speed of 420 rpm and performed in 20 mM HEPES buffer (pH 7.4), 150 mM NaCl, 5% (v/v) glycerol and

0.05% (v/v) Tween20. Binding data were analyzed using the MicroCal LLC ITC200 software. Aggregation problems due to stirring limited the upper range of concentrations of $BabA^{AD}$:Nb-ER19 in the ITC cell, such ITC curves for low affinity (low mM to high μM range) interactions between $BabA^{AD}$:Nb-ER19 variants and bg carbohydrates did not reach a plateau in the concentration range used. To allow convergence of the fitting algorithm, the N-value of the interaction fixed to 1. The unrestrained determination of N values in the high affinity ITC binding experiments between $BabA^{AD}$:Nb-ER19 variants and bg carbohydrates demonstrated the correctness of this assumption.

5.7. Leb-binding activity of H. pylori strains analyzed by Radio Immuno Assay (RIA)

The chloramine T method (Aspholm et al., 2006) was used to ^{125}I label HSA:Le^b5 conjugates (Le^b conjugate) (IsoSep AB, Tullinge, Sweden). For a standard binding assay, 1 ng of ^{125}I HSA:Le^b5 conjugate (^{125}I-Leb) was diluted with 300 ng cold Le^b conjugate for improved equilibrium in binding, and in addition for test of binding capacity. This cocktail was applied to 1 mL of bacterial cells (OD_{600} 0.1), samples were incubated for >1h to equilibrium and bacterial cells were pelleted by centrifugation. Counts per minute measurements of pellet vs. supernatant were carried out on a 2470 Wizard^2 Automatic Gamma counter (PerkinElmer, Waltham, MA, USA).

5.8. DTT and NAC titration of H. pylori and binding measurements by RIA

All RIA were obtained essentially as described above. In Figure 4A two experiments where performed. In the first: *H. pylori* 17875 was: (1) first mixed with DTT in varying concentrations for 1h followed by resuspension in DTT-neutral buffer and addition of 1 ng hot ^{125}I-Le^b conjugate mixed with 300 ng cold similar conjugate for a second hour, giving rise to the "pre-treatment" regime (dotted lines) or; (2) simultaneously mixed with both the ^{125}I-Le^b conjugate and DTT for 1 hour (black line). In the second experiment: *H. pylori* 17875 was tested for reformation of CL2 bond and recovery of binding activity by removal and reconditioning of redox conditions, i.e. removal of reducing DTT agent. First, the bacterial suspensions were mixed with varying concentrations of DTT for 30 min. Bacteria were either resuspended in DTT-neutral buffer and incubated with the 1 ng hot and 300 ng cold ^{125}I-Leb-conjugate for 2 h or left in DTT during this time, just adding the conjugate mix for subsequent RIA-analysis. Experiment 1 and 2 were normalized and used to make up Figure 4A, where data points show the mean ± s.d. (n=2). In Figure 4C, 1 mL (OD_{600}=0.1) samples of various *H. pylori* clinical isolates were RIA-tested for Leb-binding. Here samples were incubated for 1.5 h in DTT, and subsequently changed to DTT-neutral buffer and subjected to a normal RIA for 1 h. RIA binding data of the individual clinical isolates was normalized, with 0 mM DTT defined as 100% binding. Data points show the mean of two experiments. In Figure 5B, 1 mL (OD_{600}=0.1) samples of *H. pylori* strain 17875 were tested

similarly by RIA for Leb-binding upon 1 h, 37°C pretreatment with Acetylcysteine (Acetylcystein, Meda, 200 mg/mL, solution for nebulizator, pH 7) in a concentration series from 0 to 50 mg/mL to determine the inhibition activity of NAC incubation on BabA mediated Le^b binding. Data points show the mean ± s.d. (n=3).

5.9 RIA binding quantification of S831G generalist vs. S831S specialist clones

In Figure 7F a competition experiment was performed where *H. pylori* S831G and S831S clone 1 (OD_{600}=0.1) were incubated with increasing concentrations of non-radiolabeled ALe^b conjugate for 30 min and were next incubated with 1 ng radiolabeled ^{125}I-Leb-conjugate mixed with 300 ng non-radiolabeled Le^b conjugate for 17 h, pellet and supernatant were separated and cpm's counted as described above. Increasing concentrations of ALe^b titrates out 34.6% of the Le^b binding signal in the generalist clone compared to the control where no ALe^b was added. When saturated with ALeb, Le^b binding of the generalist clone becomes similar to that of the specialist clone, demonstrating the concomitant expression of the Locus A generalist and Locus B specialist *babA* variants in S831G. Values are means ± s.d.

6. Electron microscopy

Grids with *H. pylori* J166*CL2* obtained fresh from an overnight plate culture were incubated for 10 min with 10 μl of 1:100 dilution of α-BabA antibody AK253 (Yamaoka et al., 2002). This antibody was previously used to detect BabA on *H. pylori* cells in electron microscopy (Bäckström et al., 2004). After washing with 3 mL of Tris buffer (10 mM Tris, 10 mM $MgCl_2$, pH 7.2) for 5 min on a plate shaker the secondary nanogold labeled anti-rabbit antibodies (GAR 10, BBI solutions) in a 1:20 dilution was added for 10 min. Negative staining was obtained with 2 % sodium silico-tungstate. Similar results were obtained with α-BabA antibody AK277 (Odenbreit et al., 2002), although with less staining as primary antibody was used in 1:600. In both experiments the CL2 mutant was strongly stained compared to the J166Wt and the J166Δ*babA*. Electron microscopy imaging was performed at Umeå Core Facility for Electron Microscopy (UCEM), Umeå University, Sweden.

7. NAC administration to *H. pylori* infected Leb-mouse, quantification of neutrophil mucosal infiltration and *H. pylori* infection load in epithelial proximity.

All animal experiments were approved by the local ethical committee for animal studies at the Medical Faculty of Umeå University, Umeå, Sweden (Protocol number "A 23-13").

FVB/N mice transgenic for the human α-1,3/4 fucosyl-transferase gene (Falk et al., 1995) were used in this study. Only male mice of 6-8 weeks of age were selected and kept in individual cages on a 12 hour light-dark cycle with free access to food and water except the days of experimental

H. pylori challenge. Two hours prior to infectious challenge mice were starved and then oro-gastrically gavaged with 2 – 5 x 10^8 (single passaged in mice) *H. pylori* strain J166 twice a week for 2 weeks. The following 2 weeks mice (25-30 grams) were administered Acetylcystein, Meda, dissolved in drinking water, 40 mg/day/animal similar to (Huynh et al., 2004). Clinically, NAC is administered to patients as 2 -4 times a 600 mg doses/day and is rapidly cleared: $t_{1/2}$ of 2.27 h (Borgström et al., 1986). Here NAC was added to the drinking water, with much lower expected dose per drinking, therefore, relative to body weight, a higher dose of NAC (8 mg/mL for 5 mL per day) was supplemented. Animals were terminated week five; stomachs were harvested and longitudinally divided: one part was used for quantitative culturing of *H. pylori* and the other part was fixed in 4 % paraformaldehyde (PFA) and used for immunohistochemistry. Sentinel mice were routinely monitored as free from common murine pathogens and all animals were defined as specific pathogen free. To ensure transgenicity, all breeding couples were tested for the presence of the transgene as described by Falk *et al*. 1995.

To study neutrophil infiltration, five-micrometer-thick tissue sections were prepared from paraffin-embedded stomach tissue and used for immunodetection of neutrophils by the avidin-biotin-peroxidase complex method. Rat monoclonal antibody MCA771G (Serotec) detecting the murine Ly-6B.2 alloantigen was used as primary antibody. The secondary antibody was Goat anti-rat IgGAb (SC-2019; Santa Cruz Biotechnology). The sections were incubated with Vectastain Elite ABC Reagent and were visualized by a diaminobenzidine (DAB) reaction (Vector Laboratories). Images were taken with a DC300F digital camera attached to a Leica DMLB microscope. Neutrophils in each image were manually counted in three independent, equally sized fields, excluding the non-relevant tissue. The means of three fields produced the individual score for every mouse and the data presented in Figure 5D shows the means of 8 and 9 mice, for the control and NAC-treated groups respectively, with ±1SD. Group means were compared with the unpaired t-test with Welch's correction.

To quantify *H. pylori* infection, i.e. the infectious load as regard to those *H. pylori* bacterial cells that are found in close proximity to the epithelium, autopsy sections from NAC-treated and control-animals were blocked and probed with rabbit anti-*H. pylori* polyclonal antibodies (Agrisera, Sweden) and mouse monoclonal anti-Le^b antibodies (Immucor Inc. CA, USA) overnight at 4°C. After washing (0.15 M NaCl, 0.05% (v/v) Tween-20) the secondary antibodies anti-rabbit IgG Alexa-488 (Invitrogen, USA) and anti-mouse IgG APC (Jackson-Immuno, USA) were applied for 2 hours, followed by counterstaining with 4',6'-Diamidino-2-Phenylindole (DAPI) (Sigma) and mounted using Dako Fluorescent Mounting Medium (Dako North America, Inc. CA, USA). The entire section of each mouse sample was digitized at 100x magnification with Zeiss

AXIOcam MRm microscope (Carl Zeiss AB, Stockholm, Sweden) by using Tile-mode settings. Images were taken with triple-channel fluorescent recording of Alexa-488 (Ex_{max} 495 nm/Em_{max} 519 nm), APC (Ex_{max} 650 nm/Em_{max} 660 nm) and DAPI (Ex_{max} 350 nm/Em_{max} 470 nm). Fluorochromes were scanned sequentially to eliminate spectral overlap between probes. Quantitative analysis of *H. pylori* located in the epithelial lining in the corpus and antrum areas of the tissue sections was performed. Each fluorescent tile-image was imported into Image J 2.0.0–rc–14/1.49g software (http://rsbweb.nih.gov/ij/). An intensity threshold was set to discriminate the detected bacterial population from the background. The number of bacteria was calculated by dividing the total area of the bacterial signal in one section by the averaged area of a single bacterium and expressed per micrometer of corpus and antrum epithelial lining. Data presented in Figure 5C shows the bacterial counts per animal per mm of immunostained gastric epithelium; horizontal lines represents the mean value per group of control (*n*=8) and NAC-treated (*n*=9) animals with ± 1 s.d.

SUPPLEMENTARY REFERENCES

Barrozo, R.M., Cooke, C.L., Hansen, L.M., Lam, A.M., Gaddy, J.A., Johnson, E.M., Cariaga, T.A., Suarez, G., Peek, R.M., Cover, T.L., et al. (2013). Functional Plasticity in the Type IV Secretion System of Helicobacter pylori. PLoS Pathog *9*, e1003189.

Bäckström, A., Lundberg, C., Kersulyte, D., Berg, D.E., Borén, T., and Arnqvist, A. (2004). Metastability of Helicobacter pylori bab adhesin genes and dynamics in Lewis b antigen binding. Proc Natl Acad Sci USA *101*, 16923–16928.

Borgström, L., Kågedal, B., and Paulsen, O. (1986). Pharmacokinetics of N-acetylcysteine in man. Eur J Clin Pharmacol *31*, 217–222.

Chiu, J., March, P.E., Lee, R., and Tillett, D. (2004). Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic Acids Res *32*, e174.

Colbeck, J.C., Hansen, L.M., Fong, J.M., and Solnick, J.V. (2006). Genotypic profile of the outer membrane proteins BabA and BabB in clinical isolates of Helicobacter pylori. Infect Immun *74*, 4375–4378.

Cowtan, K. (1994). "dm": An automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography *31*, 34–38.

Dailidiene, D., Dailide, G., Kersulyte, D., and Berg, D.E. (2006). Contraselectable Streptomycin Susceptibility Determinant for Genetic Manipulation and Analysis of Helicobacter pylori. Appl Environ Microbiol *72*, 5908–5914.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126–2132.

Ghahroudi, M.A., Desmyter, A., Wyns, L., Hamers, R., and Muyldermans, S. (1997). Selection and

identification of single domain antibody fragments from camel heavy-chain antibodies. FEBS Letters *414*, 521–526.

Huynh, H.Q., Couper, R.T.L., Tran, C.D., Moore, L., Kelso, R., and Butler, R.N. (2004). N-acetylcysteine, a novel treatment for Helicobacter pylori infection. Dig Dis Sci *49*, 1853–1861.

Jiménez-Soto, L.F., Clausen, S., Sprenger, A., Ertl, C., and Haas, R. (2013). Dynamics of the Cag-type IV secretion system of Helicobacter pylorias studied by bacterial co-infections. Cell Microbiol n/a–n/a.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallogr *26*, 795–800.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc *3*, 1171–1179.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658–674.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr *67*, 355–367.

Odenbreit, S., Kavermann, H., Püls, J., and Haas, R. (2002). CagA tyrosine phosphorylation and interleukin-8 induction by Helicobacter pylori are independent from AlpAB, HopZ and Bab group outer membrane proteins. Int J Med Microbiol *292*, 257–266.

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D Biol Crystallogr *66*, 479–485.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235–242.

Yamaoka, Y., Souchek, J., Odenbreit, S., Haas, R., Arnqvist, A., Borén, T., Kodama, T., Osato, M.S., Gutierrez, O., Kim, J.G., et al. (2002). Discrimination between Cases of Duodenal Ulcer and Gastritis on the Basis of Putative Virulence Factors of Helicobacter pylori. J Clin Microbiol *40*, 2244–2246.