



HAL
open science

Molecular Characterization of the EhaG and UpaG Trimeric Autotransporter Proteins from Pathogenic *Escherichia coli*

M. Totsika, J. Wells, C. Beloin, J. Valle, L. P. Allsopp, N. P. King, J.-M. Ghigo, M. A. Schembri

► **To cite this version:**

M. Totsika, J. Wells, C. Beloin, J. Valle, L. P. Allsopp, et al.. Molecular Characterization of the EhaG and UpaG Trimeric Autotransporter Proteins from Pathogenic *Escherichia coli*. *Applied and Environmental Microbiology*, 2012, 78 (7), pp.2179 - 2189. 10.1128/AEM.06680-11 . pasteur-01385600

HAL Id: pasteur-01385600

<https://pasteur.hal.science/pasteur-01385600v1>

Submitted on 5 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Molecular characterisation of the EhaG and UpaG trimeric autotransporter proteins**
2 **from pathogenic *Escherichia coli***

3

4 Makrina Totsika¹, Timothy J. Wells¹, Christophe Beloin^{2,3}, Jaione Valle^{2,4}, Luke P Allsopp¹,
5 Nathan P King, Jean-Marc Ghigo^{2,3} and Mark A. Schembri^{1*}

6 ¹: *Australian Infectious Disease Research Centre, School of Chemistry and Molecular*
7 *Biosciences, University of Queensland, Brisbane QLD 4072, Australia.*

8 ²: *Institut Pasteur, Unité de Génétique des Biofilms, Département de Microbiologie F-75015*
9 *Paris, France.*

10 ³: *CNRS, URA2172, F-75015 Paris, France.*

11 ⁴: *current address, Laboratory of Microbial Biofilms, Instituto de Agrobiotecnología,*
12 *Universidad Pública de Navarra-CSIC-Gobierno de Navarra, 31006 Pamplona, Spain.*

13

14

15 Running title: EhaG and UpaG TAAs of *E. coli*.

16

17 Key words: adhesion, trimeric autotransporter, *Escherichia coli*

18

19 * Corresponding author.

20

21 Mailing address: School of Chemistry and Molecular Biosciences, Building 76, University of
22 Queensland, Brisbane QLD 4072, Australia. Phone: +617 33653306; Fax: +617 33654699;

23 E-mail: m.schembri@uq.edu.au

24

25

26 **Abstract**

27 Trimeric autotransporter proteins (TAAs) are important virulence factors of many Gram-
28 negative bacterial pathogens. A common feature of most TAAs is the ability to mediate
29 adherence to eukaryotic cells or extracellular matrix (ECM) proteins via a cell-surface
30 exposed passenger domain. Here we describe the characterization of EhaG, a TAA identified
31 from enterohaemorrhagic *E. coli* (EHEC) O157:H7. EhaG is a positional orthologue of the
32 recently characterized UpaG TAA from uropathogenic *E. coli* (UPEC). Similar to UpaG,
33 EhaG localized at the bacterial cell surface and promoted cell aggregation, biofilm formation
34 and adherence to a range of ECM proteins. However, the two orthologues display differential
35 cellular binding; EhaG mediates specific adhesion to colorectal epithelial cells while UpaG
36 promotes specific binding to bladder epithelial cells. The EhaG and UpaG TAAs contain
37 extensive sequence divergence in their respective passenger domains that could account for
38 these differences. Indeed, sequence analyses of UpaG and EhaG homologues from several *E.*
39 *coli* genomes revealed grouping of the proteins in clades almost exclusively represented by
40 distinct *E. coli* pathotypes. The expression of EhaG (in EHEC) and UpaG (in UPEC) was
41 also investigated and shown to be significantly enhanced in a *hns* isogenic mutant, suggesting
42 that H-NS acts as a negative regulator of both TAAs. Thus, while the EhaG and UpaG TAAs
43 contain some conserved binding and regulatory features, they also possess important
44 differences that correlate with the distinct pathogenic lifestyles of EHEC and UPEC.

45

46

47

48

49 **Introduction**

50 Trimeric autotransporter adhesins (TAAs) are a sub-group of AT proteins that form a stable
51 trimer on the bacterial cell surface (12). TAAs have been identified in a wide range of Gram-
52 negative bacteria and where characterized are universally associated with virulence (30).
53 TAAs are defined by the presence of a short 70-100 amino acid C-terminal membrane anchor
54 domain encoding four β -sheets that forms a trimer to create a full sized β -barrel pore (49).
55 This pore facilitates the translocation of the passenger domain to the cell surface (40, 49). This
56 feature of TAAs is different to conventional AT proteins, which possess a translocation
57 domain composed of around 300 amino acids that encode 12-14 aliphatic β -pleated sheets
58 that together form a β -barrel pore (31, 59).

59 The passenger domain of TAAs requires trimerisation for stability and adhesive activity (11).
60 Modelling of TAAs such as YadA from *Yersinia enterocolitica*, Hia from *Haemophilus*
61 *influenzae* and BadA *Bartonella henselae* has revealed three distinct regions within the
62 passenger domain; an N-terminal head, a neck and a stalk (35, 50, 60). The N-terminal head
63 structure differs between TAA proteins and is primarily involved in the adhesion properties
64 of the protein (35, 60). The YadA head structure contains single-stranded, left-handed β -
65 helices, the interface of which is formed by periodically occurring, conserved sequence
66 motifs (35). These motifs can be found in the predicted head structures of many TAAs (50).

67 The head is connected to the stalk by a short, highly conserved sequence (the neck) which
68 functions as an adapter between the large globular head and the narrow stalk domain. The
69 neck is thought to act like a 'safety pin', holding the three monomers together, partly
70 explaining the stability of trimeric proteins (35). The stalk domain is repetitive, fibrous and
71 highly divergent in length with its primary function to extend the head domain away from the
72 surface of the bacterium (30). It can however confer other functional properties such as serum
73 resistance (29, 40). The variable number of repeats in the stalk domain can lead to major
74 differences in the size of TAAs. For example, BadA is more than 3,000 amino acids in length
75 (39) while YadA is only 422 amino acids (26). Large TAAs like BadA have intermittent neck
76 domains throughout the long stalk structure (30). The domain build-up of TAA proteins has
77 been demonstrated from the crystal structure of EibD, an immunoglobulin binding TAA
78 protein from *E. coli* (29, 40).

79 Recently, a TAA from the uropathogenic *E. coli* (UPEC) strain CFT073 was identified
80 (UpaG) that mediates adhesion to human bladder epithelial cells (55). UpaG promotes cell
81 aggregation and biofilm formation on abiotic surfaces by CFT073 and various other UPEC
82 strains as well as binding to the extracellular matrix (ECM) proteins fibronectin and laminin
83 (55). Prevalence studies indicated that *upaG* is frequently associated with extra-intestinal *E.*
84 *coli* (ExPEC) strains (55). UpaG has also been identified as a potential protective antigen in
85 ExPEC (19).

86 Enterohemorrhagic *Escherichia coli* (EHEC) are a pathogenic sub-class of diarrheagenic *E.*
87 *coli* (DEC). Here we identify a TAA from *E. coli* O157:H7, EhaG, which is a positional
88 orthologue of UpaG but contains significant sequence divergence within the passenger-
89 encoding domain. Cloning and expression of the *ehaG* gene from *E. coli* EDL933 revealed
90 the EhaG TAA possesses different functional properties to UpaG. These functional properties
91 correlate with the distinct tissue tropism of EHEC and UPEC pathogens. While UpaG and
92 EhaG displayed different functional characteristics, their expression in UPEC and EHEC is
93 regulated in a common fashion by H-NS.

94

95 **Materials and Methods**

96 **Bacterial strains and growth conditions.** The following *E. coli* strains were used in this
97 study: BL21(DE3) (Stratagene), MG1655, MS427 (MG1655*flu*) (38), OS56 (MG1655*flu*
98 *gfp*⁺) (48), UPEC CFT073 (33), EHEC EDL933 (37) and CFT073*upaG* (55). Cells were
99 routinely grown at 28°C or 37°C on solid or in liquid lysogeny broth (LB) medium (5),
100 supplemented with the appropriate antibiotics; kanamycin (Kan, 100 µg/ml),
101 chloramphenicol (Cam, 30 µg/ml), ampicillin (Amp, 100 µg/ml). For growth in defined
102 conditions M63B1 supplemented with 0.4% glucose (M63B1_{Glu}) media was used as indicated
103 (42).

104 **DNA manipulations and genetic techniques.** DNA techniques were performed as
105 previously described (42). Isolation of plasmid DNA was carried out using the QIAprep Spin
106 Miniprep Kit (Qiagen). Restriction endonucleases were used according to the manufacturer's
107 specifications (New England Biolabs). Chromosomal DNA purification was made using the
108 DNeasy Blood and Tissue kit (Qiagen). Oligonucleotides were purchased from Sigma
109 (Australia or France). All polymerase chain reactions requiring proofreading were performed
110 with the Expand High Fidelity Polymerase System (Roche) as described by the manufacturer.
111 Amplified products were sequenced to ensure fidelity of the PCR. DNA sequencing was
112 performed using the ABI Big Dye ver3.1 Kit (ABI) by the Australian Equine Genetics
113 Research Centre, University of Queensland, Brisbane. Prevalence studies for the *upaG* and
114 *ehaG* genes used *Taq* DNA polymerase, as described by the manufacturer (New England
115 Biolabs), with the primers 144 (5'-aataccagagcattactaacctg) and 145 (5'-
116 accttgtaattgtagacccaa).

117 **Construction of plasmids.** The *ehaG* gene was amplified by PCR from EHEC EDL933
118 using specific primers designed from the available genome sequence (130: 5'-
119 CGCGCTCGAGATAATAAGGAacattaatgaacaaaatatttaaag and 131: 5'-
120 CGCGCAAGCTTttaccactgaataaccggcaccg). The PCR product was digested with *Xho*I
121 (forward primer) and *Hind*III/*Eco*RI (reverse primer) and ligated to *Xho*I-*Hind*III/*Eco*RI
122 digested plasmid pBAD/*Myc*-HisA. The resultant plasmid (pOMS01) was then digested with
123 *Eco*RI and ligated with a correspondingly digested kanamycin-resistance encoding gene
124 cassette to give rise to plasmid pEhaG-kan (pOMS15). Resistance to kanamycin was required
125 to facilitate transformation of this plasmid into the *flu*-negative, *gfp*-positive *E. coli* K-12

126 strain OS56. The *upaG* gene from UPEC CFT073 has been described previously (55). Both
127 genes were cloned using the same strategy, with expression of *ehaG/upaG* under control of
128 the arabinose-inducible *araBAD* promoter (24). Neither gene was cloned as a fusion to the
129 6xHis-encoding sequence of pBAD/*Myc*-HisA.

130 **Construction of mutants.** In order to mutate the *ehaG* gene in EDL933 and create a *lacZ*
131 reporter transcriptional fusion in CFT073, we used homologous recombination mediated by
132 λ -red recombinase and either a one-step PCR procedure with 50-bp homology arms for
133 recombination or a three-step PCR procedure with 500-bp homology arms for recombination
134 (7, 14, 15, 52) The primers used to disrupt the *ehaG* gene in EDL933 were 843 (5'-
135 acagctaaagagtgcaactgg), 844 (5'-ccatgagcgcgacgtatcc), 845 (5'-
136 gaagcagctccagcctacactaatgatgctcgtattccttg) and 846 (5'-
137 ctaaggaggatattcatatgtgatccattaagttagtgactaagg). The mutation was confirmed using primers
138 789 (5'-aggagcccgccataaactg) and 790 (5'-ggtttaacggttgaggacaac) and subsequent
139 sequencing. A *upaG-lacZ* reporter transcriptional fusion in CFT073 Δ *lac* was constructed
140 using the same approach but employing primers *upaG.lacZzeo.L-5*
141 (cagcttctgcgcttatatcaaggaatagagacatcaataatgacctgattacggattc) and *upaG.lacZzeo.L-3*
142 (catcaggcaatgtggcggtttaccattgttaatggatgatcagctcctgctcctcgccac). Mutants were confirmed via
143 PCR and sequencing using primers *upaG.ext-5* (aggaattcatcctatgaacc) and *upaG.ext-3*
144 (ttatcgttcgaactgctactgtc). CFT073 Δ *lac upaG::lacZ-zeo* mutants were screened after
145 mutagenesis with the suicide plasmid pSC189 carrying kanamycin-resistant Mariner
146 transposon described previously (8, 13). Sequencing of the transposon in both directions
147 enabled the identification site of insertion and, hence, the gene disrupted. Mutation of the *hns*
148 gene in EDL933 and CFT073 was performed as previously described (1).

149 **Biofilm assays.** Biofilm formation on polystyrene surfaces was monitored by using 96-well
150 microtitre plates (IWAKI) essentially as previously described (46). Briefly, cells were grown
151 for 18 h in LB (containing 0.2% arabinose for induction of AT-encoding genes) at 37°C,
152 washed to remove unbound cells, and stained with crystal violet. Quantification of bound
153 cells was performed by addition of acetone-ethanol (20:80) and measurement of the dissolved
154 crystal violet at an optical density of 590 nm. Flow chamber experiments were performed as
155 previously described (27, 45). Briefly, biofilms were allowed to form on glass surfaces in a
156 multi-channel flow system that permitted continuous monitoring of their structural
157 characteristics. Flow cells were inoculated with OD₆₀₀ = 0.02 standardized cultures pre-

158 grown overnight in M9 medium containing arabinose and kanamycin. Biofilm development
159 was monitored by confocal scanning laser microscopy at 18 h post inoculation. For analysis
160 of flow cell biofilms, z-stacks were analysed using the COMSTAT software program (25).

161 **Binding to ECM components.** Bacterial binding to ECM components was performed in a
162 microtiter plate ELISA assay essentially as previously described (55). Microtiter plates
163 (Maxisorp; Nunc) were coated overnight at 4°C with 2 µg of the following ECM proteins:
164 collagen (types I-V), fibronectin, fibrinogen, laminin, elastin, heparin sulfate, human serum
165 album, BSA (Sigma-Aldrich) or the glycoproteins *N*-acetyl-D-galactosamine (NaGal), *N*-
166 acetyl-D-glucosamine (NaGlu) or *N*-acetylneuraminic acid (NaNa). Wells were washed twice
167 with TBS (137mM NaCl, 10mM Tris, pH 7.4) and then blocked with TBS-2% milk for 1 h.
168 After being washed with TBS, 200 µl of washed and standardized (OD₆₀₀ = 0.1) cultures of
169 *E. coli* strains MS427pEhaG, MS427pUpaG or MS427pBAD was added to 12 replicate wells
170 per ECM component and the plates were incubated at 37°C for 2 h. After being washed to
171 remove non-adherent bacteria, adherent cells were fixed with 4% paraformaldehyde, washed,
172 and incubated for 1 h with anti-*E. coli* serum (Meridian Life Sciences Inc., #B65001R)
173 diluted 1:500 in 0.05% TBS-Tween, 0.2% skim milk, washed and incubated for 1 h with a
174 secondary anti-rabbit conjugated horseradish peroxidase antibody (diluted 1:1000) (Sigma-
175 Aldrich; #A6154). After a final wash adhered bacteria were detected by adding 150 µL of
176 0.3mg/ml ABTS [2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)] Sigma-Aldrich) in
177 0.1M citric acid pH 4.3, activated with 1 µl/ml 30% hydrogen peroxide] and the absorbance
178 measured at 405 nm.

179 **Epithelial cell binding assays.** T24 human bladder carcinoma cells and colorectal epithelial
180 (Caco-2) cells were purchased from the American Type Culture Collection and cultured
181 according to standard protocols. The adherence of MS427pEhaG and MS427pUpaG to
182 cultured T24 or Caco-2 cells was examined essentially as previously described (6, 55).
183 Briefly, wells were seeded with 1.5 x 10⁵ epithelial cells in 24-well tissue culture plates and
184 incubated at 37°C in 5% CO₂ until they were confluent. Overnight bacterial cultures were
185 diluted and grown to an OD₆₀₀ = 0.5 in LB with appropriate antibiotics (and 0.2% arabinose
186 for *ehaG/upaG* induction). Bacterial cells were washed and added to triplicate epithelial cell
187 monolayers at a multiplicity of infection of 10. Incubation was carried out for 1.5 h at 37°C in
188 5% CO₂. After incubation, monolayers were washed five times with PBS to remove non-
189 adherent bacteria. The remaining bacteria were then released by eukaryotic cell lysis with

190 0.1% Triton X-100. The number of adherent bacteria, as well as the inoculating dose, was
191 determined by serial dilution and plating on LB agar. All experiments were performed in
192 triplicate.

193 Co-infection of Caco-2 epithelial cell monolayers with fluorescently tagged MS427pEhaG
194 and MS427pUpaG was carried out essentially as described above, but with the following
195 modifications. Seeding of epithelial cells was performed in 4-chamber glass culture slides
196 (BD Falcon). The two bacterial strains were mixed at 1:1 ratio and incubated with the Caco-2
197 cell monolayers for 1.5 h at 37°C in 5% CO₂. Following incubation, monolayers were washed
198 five times with PBS to remove non-adherent bacteria and the adherent bacteria were fixed
199 with 4% paraformaldehyde (PFA). The PFA was then removed by washing with PBS and
200 adherent bacteria were visualised using fluorescence microscopy. Adherent bacteria were
201 counted from 40 fields of view. In these experiments, the level of EhaG and UpaG protein
202 produced by MS427pEhaG and MS427pUpaG, respectively, was similar (data not shown).

203 **Purification of 6xHistidine-tagged EhaG', antibody production and immunoblotting.** A
204 480 bp segment from the passenger-encoding domain of *ehaG* was amplified by PCR with
205 primers 1707 (5'-tacttccaatccaatgcaaacgctattgcatagtgctg) and 1708 (5'-
206 ttatccacttccaatgttaaagctgtgaaccattaatgg) from *E. coli* EDL933 genomic DNA; underlined
207 nucleotides represent the ligation-independent cloning (LIC) overhangs used for insertion
208 into plasmid pMCSG7 via LIC cloning (16). The resultant plasmid (pEhaGTruncated)
209 contained the base pairs 652-1132 of *ehaG* fused to a 6xHis encoding sequence. *E. coli* BL21
210 was transformed with plasmid pEhaGTruncated, induced with IPTG and the resultant 6xHis-
211 tagged EhaG truncated protein (containing amino acids 218-378 of EhaG) was assessed by
212 SDS-PAGE analysis as previously described (53). Polyclonal anti-EhaG serum was raised in
213 rabbits by the Institute of Medical and Veterinary Sciences (South Australia). A polyclonal
214 antiserum raised against UpaG has been described previously (55). For immunoblotting,
215 whole-cell lysates were subjected to SDS-PAGE using NuPAGE[®] Novex[®] 3-8% Tris-acetate
216 precast gels with NuPAGE[®] Tris-Acetate running buffer and subsequently transferred to
217 polyvinylidene difluoride (PVDF) microporous membrane filters using the iBlot[™] dry
218 blotting system as described by the manufacturer (Invitrogen). EhaG or UpaG rabbit
219 polyclonal antiserum was used as primary and the secondary antibody was alkaline
220 phosphatase-conjugated anti-rabbit IgG. Sigma Fast[™] 5-bromo-4-chloro-3-

221 indolylphosphate–nitroblue tetrazolium (BCIP/NBT) was used as the substrate in the
222 detection process.

223 **β -galactosidase assays.** β -galactosidase assays were performed essentially as previously
224 described (32). Briefly, strains carrying *lacZ* fusions were grown on LB plates for 16 h then
225 inoculated into M63B1_{Glu} minimal medium. After 16-18 h of growth the culture was diluted
226 in Z-buffer (60mM Na₂HPO₄, 40mM NaH₂PO₄, 50mM β -mercaptoethanol, 10mM KCl,
227 1mM MgSO₄, pH 7), 0.004% SDS and chloroform was added and the samples were vortexed
228 to permeabilize the cells. Samples were incubated at 28°C and the reaction was initiated by
229 the addition of ONPG. Reactions were stopped with the addition of sodium bicarbonate and
230 the enzymatic activity was assayed in quadruplicate for each strain by measuring the
231 absorbance at 420 nm. Where required, β -galactosidase activity was also observed on LB
232 agar plates containing 5-Bromo-4-chloro-3-indolyl β -D-galactoside (X-gal).

233 **DNA curvature prediction and electrophoretic mobility shift assays.** The *upaG* promoter
234 region was analysed *in silico* using Bendit, a program that enables the prediction of a
235 curvature-propensity plot calculated with DNase I-based parameters
236 (<http://hydra.icgeb.trieste.it/dna/>) (56). The curvature is calculated as a vector sum of
237 dinucleotide geometries (roll, tilt and twist angles) and expressed as degrees per helical turn
238 ($10.5^\circ/\text{helical turn} = 1^\circ/\text{bp}$). Experimentally tested curved motifs produce curvature values of
239 $5\text{--}25^\circ/\text{helical turn}$, whereas straight motifs give values below $5^\circ/\text{helical turn}$. The 250 bp *upaG*
240 promoter region was amplified using primers *upaG*-250-5 (ttagcaaatggcagcaatt) and
241 *upaG*+1-3 (tattgatgctctctattct) and its intrinsic curvature assessed by comparing its
242 electrophoretic mobility with that of an unbent marker fragment (Promega 100 bp DNA
243 ladder) on a 0.5% TBE, 7.5% PAGE gel at 4°C for retarded gel electrophoretic mobility.

244 Gel shift assays were performed essentially as previously described (4). A DNA mixture
245 comprising the PCR amplified *upaG* promoter region and TaqI-SspI digested pBR322 at
246 equimolar ratio was incubated at room temperature for 15 minutes with increasing amounts
247 of native purified H-NS protein (a gift from Dr S. Rimsky) in 30 μ l of reaction mixture
248 containing 40mM Hepes pH 8, 60mM potassium glutamate, 8mM magnesium aspartate,
249 5mM dithiothreitol, 10% glycerol, 0.1% octylphenoxypolyethoxyethanol, 0.1mg/ml BSA (H-
250 NS binding buffer). DNA fragments and DNA-protein complexes were resolved by gel

251 electrophoresis (0.5% TBE, 3% MS agarose gel run at 50 V at 4°C) and visualized after
252 staining with ethidium bromide.

253

254 **Results**

255 **Sequence variation in *E. coli* UpaG homologues correlates with strain pathotype.** We
256 used the translated amino acid sequence of *upaG* to probe all protein sequences encoded in
257 28 *E. coli* genomes available in the NCBI database (Table S1). An intact gene encoding a
258 putative TAA was found in 24 of the 28 genomes at the same genomic location as *upaG* in
259 CFT073. Interestingly, three of the *E. coli* strains lacking the gene were commensal strains
260 (ATCC8739, K-12 MG1655 and W3110). Enteropathogenic *E. coli* (EPEC) strain E2348/69
261 contains a truncated gene due to a frameshift mutation. Multiple alignment of the 24 putative
262 TAA protein sequences revealed that the signal sequence and translocation domain are highly
263 conserved, while the passenger domain is highly variable (Fig. 1A). The passenger domain of
264 all sequences contained Hep_Hag and HIM motifs, however diversity was observed in both
265 motif sequence and number (Fig 1A). Phylogenetic analysis of the 24 translated full-length
266 protein sequences revealed grouping according to strain pathotype, with sequences from DEC
267 strains clading separately from those encoded in ExPEC strains (Fig 1B). Proteins belonging
268 to DEC strains shared high sequence identity (95%), while proteins encoded in ExPEC
269 genomes shared 53% sequence identity, significantly higher than the overall 38% identity
270 shared among the 24 *E. coli* TAA proteins examined.

271 ***upaG* is highly prevalent among diarrheagenic *E. coli*.** We have previously demonstrated
272 that *upaG* is commonly found in UPEC isolates (55). However, the primers used in our
273 previous molecular screening were designed to amplify sequence from the passenger domain
274 of *upaG* that is not conserved among the gene homologues found in the genome of DEC
275 strains. In order to determine the prevalence of *upaG* homologues in DEC strains, we
276 designed a new set of primers specific to a highly conserved region within the translocation
277 domain and screened a collection of Shiga toxin-producing *E. coli* (STEC) isolates. We also
278 screened a collection of UPEC isolates with the new set of primers. A product of the correct
279 size was found in 92% (51/55) of STEC and 86.5% (64/74) of UPEC strains, indicating that
280 *upaG* homologues are highly prevalent among pathogenic *E. coli* of intestinal as well as
281 extraintestinal origin.

282 **Comparative sequence analysis of UpaG and EhaG from prototypic strains UPEC**
283 **CFT073 and EHEC EDL933, respectively.** We hypothesized that the pathotype-specific
284 grouping of UpaG homologues, due to the variable sequence of the function-encoding

285 passenger domain, suggests that TAA proteins may have evolved specificity for their host
286 environment. To test this tenet, we selected one representative protein from the ExPEC group
287 and one from the DEC clade to perform a detailed comparative sequence and functional
288 analysis. UpaG from the prototypic UPEC strain CFT073 was chosen as the only functionally
289 characterized TAA member from *E. coli* and was compared to its positional orthologue in the
290 prototypic EHEC strain EDL933 (gene *z5029*). The *z5029* gene was named *ehaG*, in a
291 fashion consistent with previously characterized AT proteins from EHEC (57) while still
292 acknowledging that it is a positional orthologue of *upaG* from UPEC. The *ehaG* gene
293 (4767bp) encodes a protein similar to UpaG, containing an extended N-terminal signal
294 sequence with predicted cleavage after amino acid 53, as well as an 89 amino acid C-terminal
295 translocation domain conserved among all TAA adhesins (Fig 2A). Despite being positional
296 orthologues, *upaG* and *ehaG* differ significantly in size and sequence (Fig 2A). The predicted
297 proteins are 1778 and 1588 amino acids, respectively, with the difference due to the presence
298 of 190 additional amino acids in the passenger domain of UpaG (Fig. 2A). In fact, although
299 their translocation domains are 100% identical, the two proteins share only 65% identity over
300 the passenger domain. Sequence analysis of the passenger domain of EhaG and UpaG against
301 the Pfam and TIGRFAM databases showed that the proteins contain 13 and 14 Hep_Hag
302 repeats, respectively. The majority of these 28 amino acid long repeats are found sequentially
303 at the N-terminal end of the passenger domain (Fig 2A). The head crystal structure of TAAs
304 YadA and BadA are composed of such domains (35, 50) and it is likely that this region
305 encodes the head of the EhaG and UpaG proteins (Fig 2A). This is also supported by domain
306 annotation for UpaG and EhaG proteins performed using the daTAA server (50) (data not
307 shown). EhaG and UpaG also encode 12 and 15 HIM motifs, respectively, dispersed along
308 the passenger domain. This highly conserved 24 amino acid motif is known to form the neck-
309 structure in YadA (35), a transition region between the globular head and the narrower stalk
310 domain of the protein. Alignment of the multiple Hep_Hag and HIM sequences present in
311 EhaG and UpaG (Fig 2B) revealed that HIM sequences are well conserved between the two
312 homologues, whereas the Hep_Hag sequences displayed some degree of variability (Fig 2B
313 and C).

314 **EhaG from EHEC EDL933 mediates autoaggregation and biofilm formation.** The
315 significant sequence differences in the passenger domain of EhaG and UpaG prompted us to
316 investigate whether EhaG possesses the same functional properties as UpaG. For this purpose

317 the *ehaG* gene was amplified from the chromosome of EDL933 and cloned into
318 pBAD/MycHis-A under the control of the inducible *araBAD* promoter (24). The pEhaG
319 vector was introduced into the previously described *E. coli flu* mutant strain MS427 (38) that
320 is deficient in Ag43 expression and thus lacks the ability to autoaggregate and form a biofilm.
321 Over-expression of EhaG in this background resulted in cell-cell aggregation from standing
322 overnight cultures within 20 minutes, in contrast to the MS427pBAD control (Fig 3A). We
323 also tested the ability of EhaG to mediate biofilm formation in two distinct systems. In the
324 polystyrene microtitre plate assay MS427pEhaG formed a significant biofilm compared to
325 the MS427pBAD control following induction with arabinose ($P < 0.001$; Fig 3B). To compare
326 the biofilm forming capacity of EhaG and UpaG, we used the dynamic model of a
327 continuous-flow chamber. Green fluorescent protein (GFP)-tagged MS427 cells (OS56)
328 expressing *ehaG* from plasmid pEhaG or *upaG* from plasmid pUpaG were monitored for
329 biofilm formation over 18 hours using confocal laser scanning microscopy. In contrast to the
330 OS56pBAD control, cells producing either EhaG or UpaG formed a strong biofilm across the
331 entire surface of the chamber to a depth of approximately 15 μ m (Fig 3C).

332 **EhaG and UpaG bind to extracellular matrix (ECM) proteins.** A common feature of
333 TAAs is the ability to mediate binding to components of the ECM such as collagen and
334 laminin. We therefore examined the ability of EhaG and UpaG to bind to a range of ECM
335 proteins. The MS427pEhaG and MS427pUpaG strains both displayed binding to laminin,
336 fibronectin, fibrinogen and collagen types I, II, III and V (Fig 4). In contrast, no UpaG/EhaG-
337 mediated binding was observed to type IV collagen, elastin, heparin sulfate, human serum
338 albumin, bovine serum albumin or the glycans *N*-acetyl-D-galactosamine (NaGal), *N*-acetyl-
339 D-glucosamine (NaGlu) and *N*-acetylneuraminic acid (NaNa) (Fig 4 and data not shown).
340 Thus, EhaG and UpaG mediate binding to selected ECM proteins, and the binding specificity
341 to these ECM proteins is conserved between both TAAs.

342 **EhaG and UpaG mediate differential cellular adhesion.** All TAAs characterized to date
343 play a role in adhesion. We have previously demonstrated that UpaG enhances adhesion of
344 CFT073 to T24 bladder epithelial cells (55). We therefore investigated if EhaG has a similar
345 function; however no increase was observed in the number of adherent bacteria recovered
346 from T24 bladder epithelial cell monolayers infected with EhaG-producing MS427 cells
347 compared to MS427pBAD (data not shown). Since EhaG is encoded by an enteric pathogen,
348 we therefore tested its ability to mediate binding to intestinal epithelial cells. The number of

349 adherent bacteria recovered after incubation of Caco-2 epithelial cell monolayers with
350 MS427pEhaG was significantly higher than the MS427pBAD control (Fig 5A). The ability of
351 EhaG to mediate bacterial adhesion to intestinal epithelial cells prompted us to investigate if
352 UpaG can also mediate adhesion to Caco-2 cells. MS427pUpaG failed to adhere to Caco-2
353 epithelial cell monolayers at significantly higher numbers than the MS427pBAD control (Fig
354 5A). Upon co-infection of Caco-2 epithelial cell monolayers with a mixed (1:1) population of
355 EhaG- and UpaG-producing MS427 cells that carried different fluorescent chromosomal
356 markers (green for MS427pEhaG and red for MS427pUpaG), EhaG-producing cells adhered
357 at 2.5-fold greater numbers than UpaG-producing cells (Fig 5B and C). Reverse competition
358 experiments using red MS427pEhaG and green MS427pUpaG cells produced a similar result
359 (data not shown), ruling out a potential fluorophore-dependent bias in bacterial numbers.

360 **Regulation of *upaG* expression by H-NS.** We did not detect expression of UpaG or EhaG in
361 protein samples prepared from *in vitro* cultured UPEC CFT073 and EHEC EDL933,
362 respectively, using western blotting with specific polyclonal antiserum (data not shown) (55).
363 To explore possible reasons for this lack of expression, we constructed a *lacZ* reporter fusion
364 to the chromosomal promoter of *upaG* in CFT073 (CFT073 Δ *lac_upaG::lacZ*) and subjected
365 the strain to random mariner-transposon mutagenesis in order to identify negative regulators
366 of *upaG* expression. Screening of 20,000 mutants identified two mutants with β -galactosidase
367 activity. Both mutants carried transposon insertions in different positions of the gene
368 encoding the global regulator H-NS (corresponding to amino acid positions 97 and 127 of the
369 H-NS protein). To confirm the regulatory role of H-NS in expression of *upaG*, we
370 constructed a *hns* deletion in CFT073 Δ *lac_upaG::lacZ* and subsequently complemented this
371 strain with a plasmid containing the *hns* gene downstream of the inducible *araBAD* promoter.
372 Relative β -galactosidase activity was measured in M63B1_{Glu} grown cultures of
373 CFT073 Δ *lac_upaG::lacZ_* Δ *hns* complemented with empty pBAD30 vector or pBAD30*hns*
374 in the absence or presence of 0.2% arabinose. Deletion of *hns* in CFT073 de-repressed *upaG*
375 expression by six-fold, which was reversed upon complementation with pBAD30*hns* but not
376 the pBAD30 control under arabinose induction (Fig 6A).

377 **H-NS binds to the *upaG* promoter region.** H-NS has an affinity for AT rich, intrinsically
378 curved double-stranded DNA. The 250 bp promoter region of *upaG* is 60.4% AT rich and we
379 examined its properties as well as its H-NS binding capacity using several approaches. First,
380 we generated an *in silico* curvature-propensity plot calculated with DNase I-based parameters

381 and showed that this 250 bp segment may adopt a curved conformation (Fig 6B). Next, we
382 demonstrated this curvature experimentally by examination of the PCR amplified 250 bp
383 fragment using polyacrylamide gel electrophoresis at 4°C. Using this method, which has been
384 used previously to demonstrate DNA curvature (54, 58), the 250 bp *upaG* promoter region
385 displayed a slightly retarded gel electrophoretic mobility compared to non-curved DNA
386 standards (Fig 6C). Finally, to demonstrate direct binding of H-NS to the 250 bp promoter
387 region of *upaG*, we performed electrophoretic mobility shift assays. The 250 bp PCR product
388 was mixed with *TaqI-SspI*-digested pBR322 DNA (which contains the *bla* promoter and has
389 been previously shown to be bound by H-NS), incubated with increasing concentrations of
390 purified H-NS protein and subsequently visualized by gel electrophoresis. The 250 bp *upaG*
391 promoter region and the fragment containing the *bla*-promoter were retarded in mobility by
392 the addition of 0.5 µM H-NS (Fig 6D). The pBR322 fragments not containing the *bla*-
393 promoter were not influenced by H-NS at these concentrations, indicating that H-NS binds
394 with specificity. These results suggest that H-NS binds to the regulatory region of *upaG* by
395 recognizing a DNA region within 250 bp 5' of the ATG translation start codon.

396 **Mutation of the *hns* gene results in increased expression of *upaG* and *ehaG*.** To confirm
397 that H-NS acts as a repressor of *upaG* expression, we constructed a *hns* isogenic mutant in
398 CFT073 and examined UpaG production by Western blot analysis. Loss of H-NS in CFT073
399 resulted in an increase in the production of UpaG that was detectable by western blot analysis
400 using an anti-UpaG specific serum (Fig 7A). We also constructed an isogenic EDL933 *hns*
401 mutant strain and tested for EhaG expression by western blot analysis using an anti-EhaG
402 specific serum. Similar to CFT073, deletion of *hns* in EDL933 result in increased expression
403 of the EhaG protein (Fig 7B). Taken together, these results demonstrate that H-NS negatively
404 regulates the expression of *upaG* in CFT073 and *ehaG* in EDL933.

405

406 **Discussion**

407 TAAs are an important group of virulence factors in many Gram-negative pathogens. TAAs
408 are translocated to the cell surface via the type V secretion pathway and adopt a trimeric
409 conformation in the outer membrane. Three TAAs have been characterized from pathogenic
410 *E. coli*, namely UpaG, Saa and the Eib group of proteins (36, 43, 55). Here we have
411 characterised the EhaG protein, a TAA from enterohaemorrhagic *E. coli* O157:H7. Like its

412 positional orthologue UpaG from UPEC, EhaG mediates bacterial cell aggregation, biofilm
413 formation and adherence to ECM proteins. However, in contrast to UpaG (which mediates
414 specific adherence to bladder epithelial cells), EhaG mediates specific adherence to intestinal
415 epithelial cells.

416 EhaG and UpaG both possess a classical TAA domain structure consisting of an N-terminal
417 signal sequence, a passenger domain and a C-terminal translocation domain. The
418 translocation domain is the most conserved feature of the TAA family, with the 70-100
419 amino acid C-terminal region responsible for translocation of the passenger domain to the
420 bacterial surface in all TAAs (40). We showed previously that the last 71 amino acids of
421 UpaG, corresponding to the L1- β subdomain, represent the translocation unit (55). This
422 region is completely conserved at the amino acid level in EhaG and UpaG.

423 In contrast to the translocation domain, the passenger domain of EhaG and UpaG exhibits
424 extensive amino acid sequence variability (65% identity). Further analysis of the passenger
425 domain revealed that the Hep_Hag motifs in the predicted head region were more variable
426 than the HIM motifs in the stalk and neck structures. The head region has been found to be
427 responsible for adhesion in other TAAs (30). Although the structural domains responsible for
428 the binding properties of EhaG and UpaG have not been defined, it is likely that sequence
429 differences in the head region account for the different EhaG and UpaG cell adherence
430 phenotypes observed in this study.

431 Most trimeric AT proteins characterised to-date display an adhesive activity mediating
432 bacterial interactions with either host cells or ECM proteins (3, 9, 10, 23, 41, 44). Both EhaG
433 and UpaG possessed several conserved features - they mediated cell aggregation, biofilm
434 formation and adhesion to laminin, fibronectin, fibrinogen and collagen types I, II, III and V.
435 There are several features of both proteins that could account for these observations. The
436 multiple Hep_Hag motifs in the head region, despite a degree of sequence variation, could
437 mediate some of these functions. Alternatively, these properties may be associated with
438 conserved structural features of both proteins rather than the presence of specific binding
439 domains.

440 We found that the sequence divergence in the passenger domain of EhaG and UpaG
441 corresponded strongly with diarrheagenic and extraintestinal *E. coli* pathotypes. In the case of
442 DEC strains, EhaG is highly conserved and exhibits >95% amino acid identity. In contrast,

443 the sequence of UpaG among ExPEC strains was more variable. The UpaG protein from
444 UPEC CFT073 has previously been shown to mediate cell aggregation, biofilm formation
445 and adhesion to the ECM proteins laminin and fibronectin (55). The extensive sequence
446 variation of UpaG suggests that there may be differences in these functional properties
447 between UpaG variants from different UPEC/ExPEC strains. Thus, a more detailed analysis
448 of the functions of different UpaG variants is required to properly assess its role in virulence.

449 Adherence of EHEC to the intestinal epithelium is essential for initiation of infection. The
450 adherence of *E. coli* O157:H7 to Caco-2 cells has been shown to proceed in two stages; an
451 initial diffuse adherence to epithelial cells followed by proliferation to develop micro-
452 colonies and intimate adherence (51). The intimate attachment and micro-colony formation
453 are thought to be mediated primarily by intimin-Tir interactions whereas the diffuse
454 adherence to Caco-2 cells requires multiple factors encoded by the LEE and additional
455 chromosomal loci (51). Some of the proteins implicated in this initial adherence include
456 ToxB and EspA (20, 51). The diffuse adherence pattern mediated by EhaG to Caco-2 cells
457 following expression in *E. coli* K-12 suggests it may contribute to initial adherence of *E. coli*
458 O157:H7 to intestinal epithelial cells. This remains to be demonstrated.

459 Immunodetection employing antibodies specific to EhaG and UpaG failed to detect these
460 proteins in whole cell extracts prepared from wild-type strains EDL933 and CFT073,
461 respectively. We were also unable to detect expression of EhaG or UpaG from wild-type
462 bacteria during interaction with cultured epithelial cells (data not shown). This prompted us
463 to examine possible mechanisms by which the expression of *ehaG* and *upaG* could be
464 repressed. We initially focused our analysis on *upaG* and identified a role for H-NS in its
465 regulation. At the transcriptional level, a significant increase in *upaG* promoter activity was
466 observed in a CFT073 Δ *lac_upaG::lacZ-zeo hns* mutant. Consistent with this result, we also
467 observed an increase in the expression of the UpaG protein by CFT073 in a *hns* mutant
468 background. H-NS is a histone-like DNA-binding protein that shows affinity for A-T rich and
469 bent nucleation sites on DNA (17, 18). Our data demonstrates that H-NS acts as a repressor
470 of *upaG* transcription, most likely through direct binding to a region comprising the 250 bp
471 upstream of the *upaG* open-reading frame. This sequence shares 83% nucleotide sequence
472 conservation with the corresponding region in *ehaG*; indeed we also observed an increase in
473 the expression of EhaG by EDL933 in a *hns* mutant background. Thus, the transcription of
474 *ehaG* and *upaG* is negatively regulated by H-NS. It is possible that the transcription of *ehaG*

475 (in EHEC) and *upaG* (in UPEC) is coordinated with other H-NS repressed genes. For
476 example, mutation of the *hns* gene in UPEC strain 536 results in the de-repression of multiple
477 virulence factors including alpha-hemolysin, iron uptake systems and fimbriae (34), and
478 several cryptic *E. coli* chaperone-usher fimbrial genes have also been shown to be repressed
479 by H-NS in *E. coli* (28). The binding of H-NS to DNA is also modulated by temperature,
480 with relief of repression for many genes observed above a threshold temperature of 32°C
481 (18). Although we did not observe expression of EhaG/UpaG at 37°C in our experiments, it is
482 possible that these proteins are expressed under specific conditions such as during host
483 infection.

484 In this study we have examined the functional properties of the trimeric autotransporter EhaG
485 from EHEC, and compared its characteristics to UpaG from UPEC. Both proteins share
486 several conserved features, yet also possess unique properties that may be associated with
487 host tissue tropism and pathogenesis. One property common to EhaG and UpaG is the ability
488 to mediate biofilm formation. The role of biofilm formation in chronic bladder infection by
489 UPEC (2) and environmental contamination of food by EHEC (47) has been documented.
490 Indeed, given the recent outbreak of the STEC O104 strain in Germany, it will be important
491 to thoroughly characterize the role of proteins such as EhaG in biofilm growth by pathogenic
492 *E. coli*.

493

494 **Acknowledgements:**

495 We thank Dr Sylvie Rimsky for providing the purified native H-NS protein. This work was
496 supported by grants from the Australian National Health and Medical Research Council
497 (631654), the Australian Research Council (DP1097032), the University of Queensland (ECR
498 grant to MT), the Institut Pasteur, the CNRS URA 2172, the Network of Excellence
499 EuroPathoGenomics, the European Community (LSHB-CT-2005-512061). MAS is
500 supported by an ARC Future Fellowship (FT100100662) and JV was a Marie-Curie Fellow.

501

502

503

504 **References**

- 505 1. **Allsopp, L. P., C. Beloin, G. C. Ulett, J. Valle, M. Totsika, O. Sherlock, J. M.**
506 **Ghigo, and M. A. Schembri.** 2011. Molecular Characterization of UpaB and UpaC -
507 two new Autotransporter Proteins of Uropathogenic *Escherichia coli* CFT073. *Infect*
508 *Immun.*
- 509 2. **Anderson, G. G., J. J. Palermo, J. D. Schilling, R. Roth, J. Heuser, and S. J.**
510 **Hultgren.** 2003. Intracellular bacterial biofilm-like pods in urinary tract infections.
511 *Science* **301**:105-107.
- 512 3. **Barenkamp, S. J.** 1996. Immunization with high-molecular-weight adhesion proteins
513 of nontypeable *Haemophilus influenzae* modifies experimental otitis media in
514 chinchillas. *Infect Immun* **64**:1246-1251.
- 515 4. **Beloin, C., and C. J. Dorman.** 2003. An extended role for the nucleoid structuring
516 protein H-NS in the virulence gene regulatory cascade of *Shigella flexneri*. *Mol*
517 *Microbiol* **47**:825-838.
- 518 5. **Bertani, G.** 1951. Studies on lysogenesis. I. The mode of phage liberation by
519 lysogenic *Escherichia coli*. *J Bacteriol* **62**:293-300.
- 520 6. **Bokil, N. J., M. Totsika, A. J. Carey, K. J. Stacey, V. Hancock, B. M. Saunders,**
521 **T. Ravasi, G. C. Ulett, M. A. Schembri, and M. J. Sweet.** 2011. Intramacrophage
522 survival of uropathogenic *Escherichia coli*: differences between diverse clinical
523 isolates and between mouse and human macrophages. *Immunobiology* **216**:1164-
524 1171.
- 525 7. **Chaveroche, M. K., J. M. Ghigo, and C. d'Enfert.** 2000. A rapid method for
526 efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic*
527 *Acids Res* **28**:E97.
- 528 8. **Chiang, S. L., and E. J. Rubin.** 2002. Construction of a mariner-based transposon
529 for epitope-tagging and genomic targeting. *Gene* **296**:179-185.
- 530 9. **Comanducci, M., S. Bambini, B. Brunelli, J. Adu-Bobie, B. Arico, B. Capecci,**
531 **M. M. Giuliani, V. Masignani, L. Santini, S. Savino, D. M. Granoff, D. A.**
532 **Caugant, M. Pizza, R. Rappuoli, and M. Mora.** 2002. NadA, a novel vaccine
533 candidate of *Neisseria meningitidis*. *J Exp Med* **195**:1445-1454.
- 534 10. **Cope, L. D., E. R. Lafontaine, C. A. Slaughter, C. A. Hasemann, Jr., C. Aebi, F.**
535 **W. Henderson, G. H. McCracken, Jr., and E. J. Hansen.** 1999. Characterization of
536 the *Moraxella catarrhalis* *uspA1* and *uspA2* genes and their encoded products. *J*
537 *Bacteriol* **181**:4026-4034.
- 538 11. **Cotter, S. E., N. K. Surana, S. Grass, and J. W. St Geme, 3rd.** 2006. Trimeric
539 autotransporters require trimerization of the passenger domain for stability and
540 adhesive activity. *J Bacteriol* **188**:5400-5407.
- 541 12. **Cotter, S. E., N. K. Surana, and J. W. St Geme, 3rd.** 2005. Trimeric
542 autotransporters: a distinct subfamily of autotransporter proteins. *Trends Microbiol*
543 **13**:199-205.
- 544 13. **Da Re, S., and J. M. Ghigo.** 2006. A CsgD-independent pathway for cellulose
545 production and biofilm formation in *Escherichia coli*. *J Bacteriol* **188**:3073-3087.
- 546 14. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal
547 genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*
548 **97**:6640-6645.
- 549 15. **Derbise, A., B. Lesic, D. Dacheux, J. M. Ghigo, and E. Carniel.** 2003. A rapid and
550 simple method for inactivating chromosomal genes in *Yersinia*. *FEMS Immunol Med*
551 *Microbiol* **38**:113-116.

- 552 16. **Donnelly, M. I., M. Zhou, C. S. Millard, S. Clancy, L. Stols, W. H. Eschenfeldt,**
553 **F. R. Collart, and A. Joachimiak.** 2006. An expression vector tailored for large-
554 scale, high-throughput purification of recombinant proteins. *Protein Expr Purif*
555 **47:446-454.**
- 556 17. **Dorman, C. J.** 2007. H-NS, the genome sentinel. *Nat Rev Microbiol* **5:157-161.**
- 557 18. **Dorman, C. J.** 2004. H-NS: a universal regulator for a dynamic genome. *Nat Rev*
558 *Microbiol* **2:391-400.**
- 559 19. **Durant, L., A. Metais, C. Soulama-Mouze, J. M. Genevard, X. Nassif, and S.**
560 **Escaich.** 2007. Identification of candidates for a subunit vaccine against
561 extraintestinal pathogenic *Escherichia coli*. *Infect Immun* **75:1916-1925.**
- 562 20. **Ebel, F., T. Podzadel, M. Rohde, A. U. Kresse, S. Kramer, C. Deibel, C. A.**
563 **Guzman, and T. Chakraborty.** 1998. Initial binding of Shiga toxin-producing
564 *Escherichia coli* to host cells and subsequent induction of actin rearrangements
565 depend on filamentous EspA-containing surface appendages. *Mol Microbiol* **30:147-**
566 **161.**
- 567 21. **Felsenstein, J.** 2005. PHYLIP (Phylogeny Inference Package) version 3.6.
568 Distributed by the author. Department of Genome Science, University of Washington,
569 Seattle.
- 570 22. **Finn, R. D., J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin,**
571 **P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. L. Sonnhammer, S. R.**
572 **Eddy, and A. Bateman.** 2010. The Pfam protein families database. *Nucleic Acids*
573 *Research* **38:D211-D222.**
- 574 23. **Geme, J. W., 3rd, and D. Cutter.** 1995. Evidence that surface fibrils expressed by
575 *Haemophilus influenzae* type b promote attachment to human epithelial cells. *Mol*
576 *Microbiol* **15:77-85.**
- 577 24. **Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith.** 1995. Tight regulation,
578 modulation, and high-level expression by vectors containing the arabinose PBAD
579 promoter. *J Bacteriol* **177:4121-4130.**
- 580 25. **Heydorn, A., A. T. Nielsen, M. Hentzer, C. Sternberg, M. Givskov, B. K. Ersboll,**
581 **and S. Molin.** 2000. Quantification of biofilm structures by the novel computer
582 program COMSTAT. *Microbiology* **146 (Pt 10):2395-2407.**
- 583 26. **Hoicyk, E., A. Roggenkamp, M. Reichenbecher, A. Lupas, and J. Heesemann.**
584 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal
585 a novel class of adhesins. *Embo J* **19:5989-5999.**
- 586 27. **Kjaergaard, K., M. A. Schembri, C. Ramos, S. Molin, and P. Klemm.** 2000.
587 Antigen 43 facilitates formation of multispecies biofilms. *Environ Microbiol* **2:695-**
588 **702.**
- 589 28. **Korea, C. G., R. Badouraly, M. C. Prevost, J. M. Ghigo, and C. Beloin.** 2010.
590 *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher
591 fimbriae with distinct surface specificities. *Environ Microbiol* **12:1957-1977.**
- 592 29. **Leo, J. C., A. Lyskowski, K. Hattula, M. D. Hartmann, H. Schwarz, S. J.**
593 **Butcher, D. Linke, A. N. Lupas, and A. Goldman.** 2011. The Structure of *E. coli*
594 IgG-Binding Protein D Suggests a General Model for Bending and Binding in
595 Trimeric Autotransporter Adhesins. *Structure* **19:1021-1030.**
- 596 30. **Linke, D., T. Riess, I. B. Autenrieth, A. Lupas, and V. A. Kempf.** 2006. Trimeric
597 autotransporter adhesins: variable structure, common function. *Trends Microbiol*
598 **14:264-270.**
- 599 31. **Loveless, B. J., and M. H. Saier, Jr.** 1997. A novel family of channel-forming,
600 autotransporting, bacterial virulence factors. *Mol Membr Biol* **14:113-123.**

- 601 32. **Miller, J. H.** 1992. A Short Course in Bacterial Genetics: A Laboratory Manual and
602 Handbook for *Escherichia coli* and Related Bacteria. Cold Spring Harbor, NY, USA:
603 Cold Spring Harbor Laboratory Press.
- 604 33. **Mobley, H. L., D. M. Green, A. L. Trifillis, D. E. Johnson, G. R. Chippendale, C.**
605 **V. Lockett, B. D. Jones, and J. W. Warren.** 1990. Pyelonephritogenic *Escherichia*
606 *coli* and killing of cultured human renal proximal tubular epithelial cells: role of
607 hemolysin in some strains. *Infect Immun* **58**:1281-1289.
- 608 34. **Muller, C. M., U. Dobrindt, G. Nagy, L. Emody, B. E. Uhlin, and J. Hacker.**
609 2006. Role of histone-like proteins H-NS and StpA in expression of virulence
610 determinants of uropathogenic *Escherichia coli*. *J Bacteriol* **188**:5428-5438.
- 611 35. **Nummelin, H., M. C. Merckel, Y. el Tahir, P. Ollikka, M. Skurnik, and A.**
612 **Goldman.** 2003. Structural studies of *Yersinia* adhesin YadA. *Adv Exp Med Biol*
613 **529**:85-88.
- 614 36. **Paton, A. W., P. Srimanote, M. C. Woodrow, and J. C. Paton.** 2001.
615 Characterization of Saa, a novel autoagglutinating adhesin produced by locus of
616 enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are
617 virulent for humans. *Infect Immun* **69**:6999-7009.
- 618 37. **Perna, N. T., G. Plunkett, 3rd, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G.**
619 **F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S.**
620 **Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T.**
621 **Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen,**
622 **D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of
623 enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**:529-533.
- 624 38. **Reisner, A., J. A. Haagenen, M. A. Schembri, E. L. Zechner, and S. Molin.** 2003.
625 Development and maturation of *Escherichia coli* K-12 biofilms. *Mol Microbiol*
626 **48**:933-946.
- 627 39. **Riess, T., S. G. Andersson, A. Lupas, M. Schaller, A. Schafer, P. Kyme, J.**
628 **Martin, J. H. Walzlein, U. Ehehalt, H. Lindroos, M. Schirle, A. Nordheim, I. B.**
629 **Autenrieth, and V. A. Kempf.** 2004. Bartonella adhesin a mediates a proangiogenic
630 host cell response. *J Exp Med* **200**:1267-1278.
- 631 40. **Roggenkamp, A., N. Ackermann, C. A. Jacobi, K. Truelzsch, H. Hoffmann, and**
632 **J. Heesemann.** 2003. Molecular analysis of transport and oligomerization of the
633 *Yersinia enterocolitica* adhesin YadA. *J Bacteriol* **185**:3735-3744.
- 634 41. **Roggenkamp, A., H. R. Neuberger, A. Flugel, T. Schmoll, and J. Heesemann.**
635 1995. Substitution of two histidine residues in YadA protein of *Yersinia enterocolitica*
636 abrogates collagen binding, cell adherence and mouse virulence. *Mol Microbiol*
637 **16**:1207-1219.
- 638 42. **Sambrook, J., E. F. Fritsch, and M. T.** 1989. Molecular cloning: a laboratory
639 manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- 640 43. **Sandt, C. H., Y. D. Wang, R. A. Wilson, and C. W. Hill.** 1997. *Escherichia coli*
641 strains with nonimmune immunoglobulin-binding activity. *Infect Immun* **65**:4572-
642 4579.
- 643 44. **Scarselli, M., D. Serruto, P. Montanari, B. Capecchi, J. Adu-Bobie, D. Veggi, R.**
644 **Rappuoli, M. Pizza, and B. Arico.** 2006. *Neisseria meningitidis* NhhA is a
645 multifunctional trimeric autotransporter adhesin. *Mol Microbiol* **61**:631-644.
- 646 45. **Schembri, M. A., K. Kjaergaard, and P. Klemm.** 2003. Global gene expression in
647 *Escherichia coli* biofilms. *Mol Microbiol* **48**:253-267.

- 648 46. **Schembri, M. A., and P. Klemm.** 2001. Biofilm formation in a hydrodynamic
649 environment by novel fimh variants and ramifications for virulence. *Infect Immun*
650 **69**:1322-1328.
- 651 47. **Seo, K. H., and J. F. Frank.** 1999. Attachment of *Escherichia coli* O157:H7 to
652 lettuce leaf surface and bacterial viability in response to chlorine treatment as
653 demonstrated by using confocal scanning laser microscopy. *J Food Prot* **62**:3-9.
- 654 48. **Sherlock, O., M. A. Schembri, A. Reisner, and P. Klemm.** 2004. Novel roles for
655 the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm
656 formation. *J Bacteriol* **186**:8058-8065.
- 657 49. **Surana, N. K., D. Cutter, S. J. Barenkamp, and J. W. St Geme, 3rd.** 2004. The
658 *Haemophilus influenzae* Hia autotransporter contains an unusually short trimeric
659 translocator domain. *J Biol Chem* **279**:14679-14685.
- 660 50. **Szczesny, P., and A. Lupas.** 2008. Domain annotation of trimeric autotransporter
661 adhesins--daTAA. *Bioinformatics* **24**:1251-1256.
- 662 51. **Tatsuno, I., H. Kimura, A. Okutani, K. Kanamaru, H. Abe, S. Nagai, K. Makino,**
663 **H. Shinagawa, M. Yoshida, K. Sato, J. Nakamoto, T. Tobe, and C. Sasakawa.**
664 2000. Isolation and characterization of mini-Tn5Km2 insertion mutants of
665 enterohemorrhagic *Escherichia coli* O157:H7 deficient in adherence to Caco-2 cells.
666 *Infect Immun* **68**:5943-5952.
- 667 52. **Ulett, G. C., J. Valle, C. Beloin, O. Sherlock, J. M. Ghigo, and M. A. Schembri.**
668 2007. Functional analysis of antigen 43 in uropathogenic *Escherichia coli* reveals a
669 role in long-term persistence in the urinary tract. *Infect Immun* **75**:3233-3244.
- 670 53. **Ulett, G. C., R. I. Webb, and M. A. Schembri.** 2006. Antigen-43-mediated
671 autoaggregation impairs motility in *Escherichia coli*. *Microbiology* **152**:2101-2110.
- 672 54. **Ussery, D. W., C. F. Higgins, and A. Bolshoy.** 1999. Environmental influences on
673 DNA curvature. *J Biomol Struct Dyn* **16**:811-823.
- 674 55. **Valle, J., A. N. Mabbett, G. C. Ulett, A. Toledo-Arana, K. Wecker, M. Totsika,**
675 **M. A. Schembri, J. M. Ghigo, and C. Beloin.** 2008. UpaG, a new member of the
676 trimeric autotransporter family of adhesins in uropathogenic *Escherichia coli*. *J*
677 *Bacteriol* **190**:4147-4161.
- 678 56. **Vlahovicek, K., L. Kajan, and S. Pongor.** 2003. DNA analysis servers: plot.it,
679 bend.it, model.it and IS. *Nucleic Acids Res* **31**:3686-3687.
- 680 57. **Wells, T. J., J. J. Tree, G. C. Ulett, and M. A. Schembri.** 2007. Autotransporter
681 proteins: novel targets at the bacterial cell surface. *FEMS Microbiol Lett* **274**:163-
682 172.
- 683 58. **Yamada, H., S. Muramatsu, and T. Mizuno.** 1990. An *Escherichia coli* protein that
684 preferentially binds to sharply curved DNA. *J Biochem* **108**:420-425.
- 685 59. **Yen, M. R., C. R. Peabody, S. M. Partovi, Y. Zhai, Y. H. Tseng, and M. H. Saier.**
686 2002. Protein-translocating outer membrane porins of Gram-negative bacteria.
687 *Biochim Biophys Acta* **1562**:6-31.
- 688 60. **Yeo, H. J., S. E. Cotter, S. Laarmann, T. Juehne, J. W. St Geme, 3rd, and G.**
689 **Waksman.** 2004. Structural basis for host recognition by the *Haemophilus influenzae*
690 Hia autotransporter. *Embo J* **23**:1245-1256.

691

692

693

695 **Figure legends**

696 **Fig 1.** Sequence analysis of UpaG homologues from 24 *E. coli* genomes available on the
 697 NCBI database and listed in Table S1. (A) Similarity plot of 24 UpaG translated amino acid
 698 sequences aligned using Clustal W. High sequence conservation (indicated by shading on the
 699 y-axis) was observed in the signal peptide (red) and translocator domain (yellow), while
 700 diversity in sequence and size was seen in the passenger domain (blue) of the proteins. (B)
 701 Unrooted phylogram of 24 UpaG translated amino acid sequences. Branch confidence levels
 702 are >90% and were determined from 1,000 bootstrap replicates of Neighbor-Joining trees
 703 calculated using PHYLIP (21). Taxon IDs represent the locus tags assigned to *upaG*
 704 orthologues in each *E. coli* genome sequence (see Table S1). Sequences from commensal and
 705 laboratory *E. coli* strains are indicated in blue, from diarrheagenic *E. coli* (DEC) in red and
 706 from extraintestinal pathogenic *E. coli* (ExPEC) in green.

707 **Fig 2.** *In silico* analysis of UpaG and EhaG homologues. (A) Schematic illustration of the
 708 domain organisation of UpaG and EhaG proteins. Indicated are the signal peptide (S.P.),
 709 translocator domain (Pfam: YadA), and the localization of the multiple invasins (Pfam:
 710 Hep_Hag) and hemagglutinin (Pfam: HIM) motifs within the passenger domain. (B) Multiple
 711 sequence alignment of the Hep_Hag repeats and HIM motifs of EhaG and UpaG. The
 712 domains were identified by sequence searches against the Pfam (22) and TIGRFAM
 713 databases hosted by the Comprehensive Microbial Resource (<http://cmr.jcvi.org>). Highly
 714 conserved residues are indicated in red. (C) Consensus sequence of the EhaG and UpaG
 715 Hep_Hag repeats and HIM motifs. Dots, no consensus; lowercase, some consensus;
 716 uppercase, high consensus.

717 **Fig 3.** EhaG promotes autoaggregation and biofilm formation in *E. coli*. (A) Settling profile
 718 of liquid suspensions of *E. coli* strains MS427pBAD (vector control) and MS427pEhaG.
 719 Suspensions were prepared from overnight LB cultures supplemented with 0.2% arabinose
 720 normalized at an OD_{600nm} of 1. Bacterial autoaggregation is inversely proportional to the
 721 optical density of each suspension measured at 600nm over a period of 60 minutes. (B)
 722 Biofilm formation by *E. coli* strains MS427pBAD (vector control) and MS427pEhaG
 723 following 18 hrs culture in LB medium supplemented with 0.2% arabinose. Biofilm
 724 formation was examined in polystyrene 96-well microtitre plates using crystal violet staining.

725 Bar charts show average absorbance measurements at 590nm \pm SEM from three independent
726 experiments. (C) Fluorescent micrographs of continuous flow biofilms formed in glass
727 chambers 18 h after inoculation with GFP-labelled *E. coli* strains (i) OS56pBAD (vector
728 control), (ii) OS56pEhaG, and (iii) OS56pUpaG. Micrographs show representative horizontal
729 sections collected within each biofilm. Shown to the right and below of each individual panel
730 are vertical sections representing the yz-plane and the xz-plane, respectively, at the positions
731 indicated by the red and green lines.

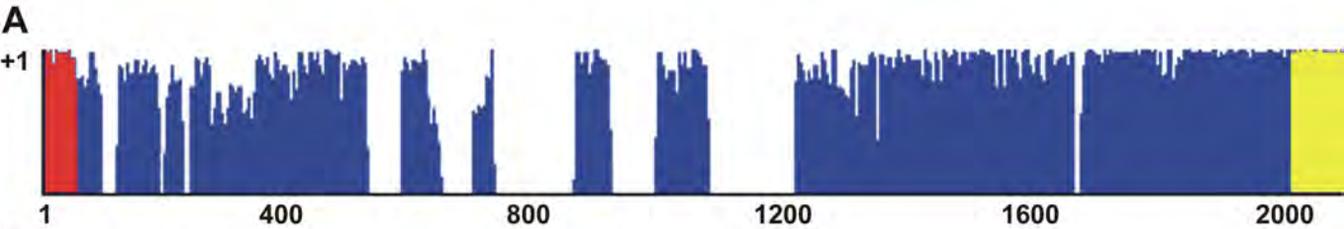
732 **Fig 4.** EhaG and UpaG mediate *E. coli* adherence to ECM proteins. ELISA-based assay
733 demonstrating binding of *E. coli* MS427pEhaG (black bars), *E. coli* MS427pUpaG (grey
734 bars), and *E. coli* MS427pBAD (white bars) to collagen (I-V), fibronectin, fibrinogen and
735 BSA. Results represent average absorbance readings at 405 nm + SEM from three
736 independent experiments. The expression of EhaG and UpaG was induced with 0.2%
737 arabinose.

738 **Fig 5.** EhaG mediates *E. coli* adhesion to intestinal epithelial cells. (A) Binding efficiency of
739 *E. coli* MS427 containing pBAD, pEhaG or pUpaG to Caco-2 epithelial cell monolayers.
740 Bars represent the average number of adherent bacteria (CFU) per epithelial cell monolayer \pm
741 SEM. (B) Competitive adhesion to Caco-2 epithelial cell monolayers of mixed (1:1) bacterial
742 inocula containing MS427green-pEhaG and MS427red-pUpaG. Bars represent average
743 number of each bacterial strain (CFU) per field of view \pm SEM. (C) Representative field of
744 view of micrographs used to quantify the number of UpaG- and EhaG-producing adherent
745 bacteria in the competitive Caco-2 epithelial cell adhesion assay. MS427pEhaG cells are
746 tagged with GFP and MS427pUpaG cells are tagged with RFP. The expression of EhaG and
747 UpaG was induced with 0.2% arabinose.

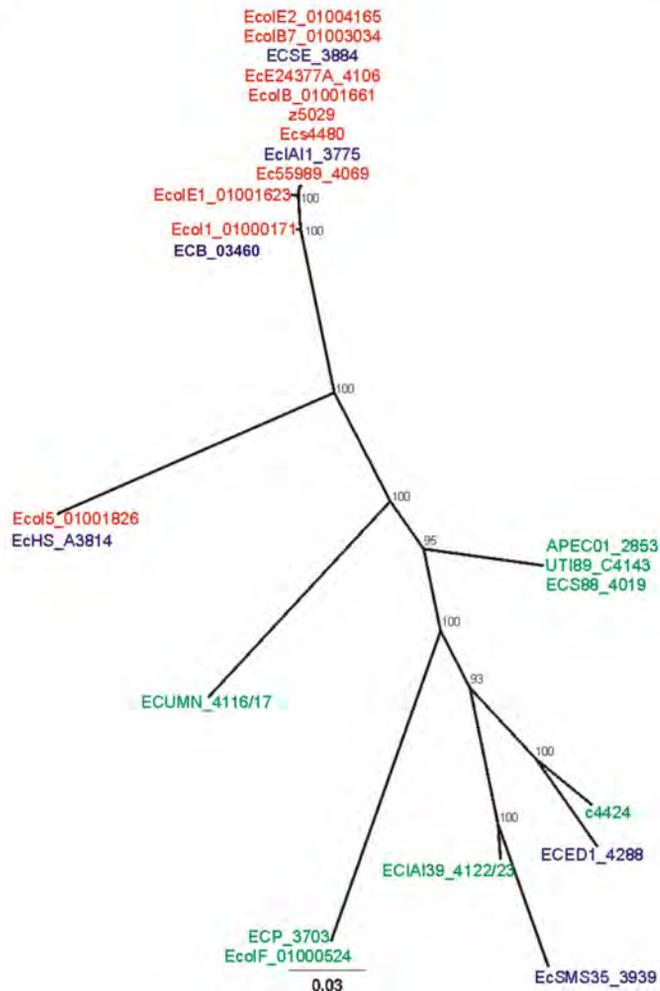
748 **Fig 6.** H-NS repression of *upaG* expression in UPEC CFT073. (A) β -galactosidase activity of
749 a chromosomal *upaG::lacZ* reporter fusion in CFT073 Δ *lac* and CFT073 Δ *lac_hns* alone, with
750 pBAD30 (vector control) or pBAD30*hns* (complementation vector). pBAD30 and
751 pBAD30*hns* containing strains were grown in presence of 0.2% arabinose. (B) Curvature plot
752 of the *upaG* promoter sequence. (C) Electrophoretic mobility of the *upaG* promoter sequence.
753 (D) Electrophoretic mobility shift assay of H-NS binding to the *upaG* and *bla* promoter
754 sequences.

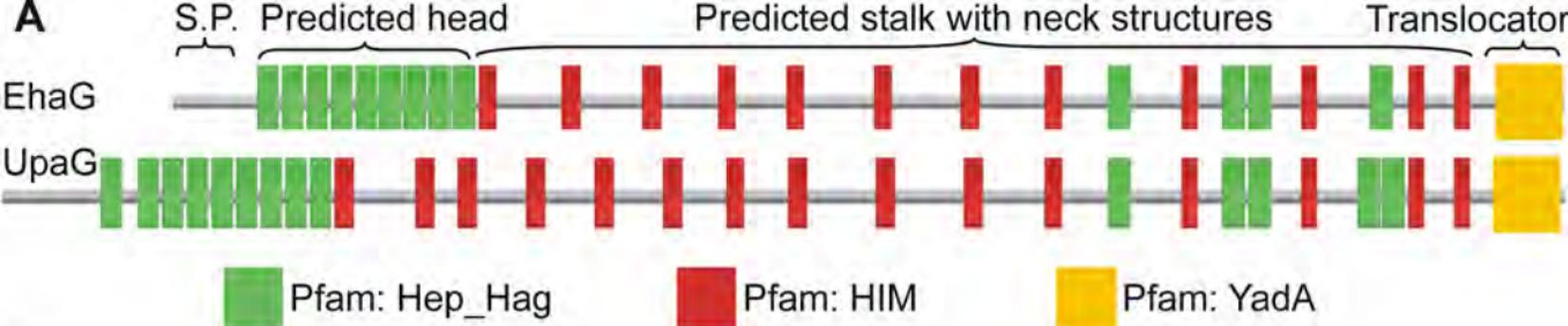
755 **Fig 7.** Production of UpaG (in CFT073) and EhaG (in EDL933) in a *hns* mutant background.
756 Western blot analysis of whole cell lysates prepared (A) from *E. coli* CFT073 and
757 CFT073*hns* employing a rabbit polyclonal anti-UpaG serum, and (B) from *E. coli* EDL933,
758 and EDL933*hns* using a rabbit polyclonal anti-EhaG serum. The monomeric UpaG and EhaG
759 proteins are indicated; possible oligomeric forms of each protein are indicated by asterisks.

760



B





B

EhaG Hep_Hag	UpaG Hep_Hag	EhaG HIM	UpaG HIM
AEQYSSAIGSKTHAIGGASMAFGVSAI	ATGGASMAFGVSAKAMGDRSVALGASSV	RKIVNVKNGAIKSDSYDAINGSQL	RKIVNMAAGAISNTSTDAINGSQL
SEGRSIALGASSYSLGQYSMALGRYSK	TNGFTSLAIGDSSLADGEKTIALGNLAK	SVITDVADGTISASSKDAVNGSQL	NKITNVAKGTVSATSTDVVNGSQL
ALGKLSIAMGDSSKAEGANAIALGNATK	AYEIMSIALGDNANASKEYAMALGASSK	SKITNVKADDLTADSTDAVNGSQL	SKITNVTAGNLTAGSTDAVNGSQL
ATEIMSIALGDTANASKAYSMAFGASSV	AGGADSLAFGRKSTANSTGSLAIGADSS	SKITNVKGDLLTGGSTDAVNGSQL	SKITNVTAGNLTAGSTDAVNGSQL
ASEENAI AIGAETEA-AENATAIGNNAK	SSNDNAI AIGNKTQALGVNSMALGNASQ	SKITNILDGTVTATSSDAINGSQL	SKITNVTAGNLTAGSTDAVNGSQL
AKGTNSMAMGFGSLADKVNTIALGNNGSQ	ASGESSIALGNTSEASEQNAIALGQGS	SVITDVADGEISDSSSDAVNGSQL	SKITNVKAGDLTAGSTDAVNGSQL
ALADNAI AIGQGNKADGVDAIALGNNGSQ	ASKVNSIALGNSLSLSEGENAIALGEGSA	SVITNVANGAISAASSDAINGSQL	SKITNVKAGDLTAGSTDAVNGSQL
SRGLNTIALGTASNATGDKSLALGNSNS	AGGSNSLAFGSQSRANGNSVAIGVGAA	SIITNVANGSISEDSTDAVNGSQL	SKITNLLAGKISSNSTDAINGSQL
ANGINSVALGADSIADLDNTVSVGNSSL	AATDNSVAIGAGSTTASNTVSVGNSAT	RQIINVADG---SEAHDAVTVRQL	SVITDVANGAVSSTSSDAINGSQL
AQGVGATAIGYNSVAKGDSSVAIGQGSY	ASGIGATAVGYNAVASHASSVAIGQDSI	RQITNVAAAG---SADTDAVNVAQL	SVITNVANGAVSATSNDAINSQL
AVGTDSLAMGAKTIVNGDKGIGIGYGAY	AVGEDSLAMGAKTIVNGNAGIGIGLNTL	RRITNVAAAG---KNATDAVNVAQL	SKITNVAAAGDLSTTSTDAVNGSQL
ANALNGI AIGSNAQVIHVNSIAIGNGST	ADAINGI AIGSNARANHADSIAMGNNGSQ	TRISNVSAG---VNNNDVVNYAQL	RQIINVADG---SEAHDAVTVRQL
AQGKDSVAIGSGSIAAADNSVALGTGSV	AAADNSVALGTGSVADEENTISVGSSTN		RQITNVAAAG---SADTDAVNVAQL
	TKYFKTNTDGADANAQKDSVAIGSGSI		RRITNVAAAG---VNATDAVNVAQL
			TRISNVSAG---VNNNDVVNYAQL

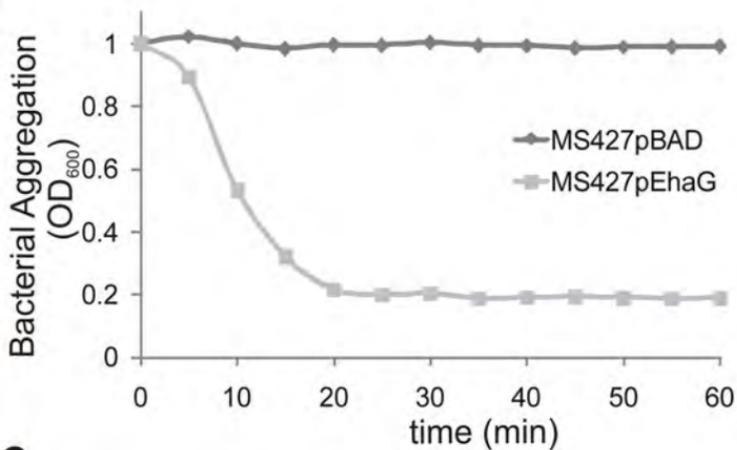
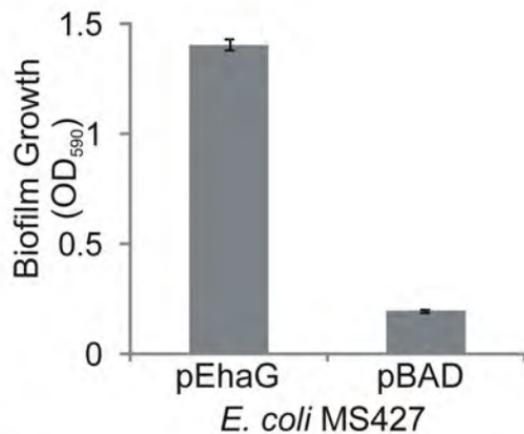
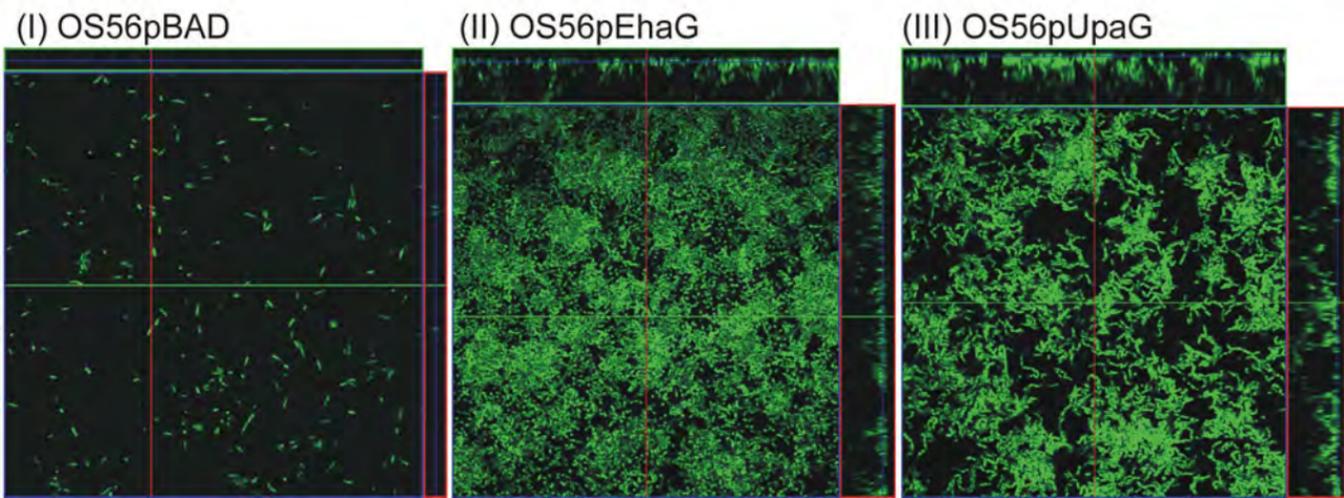
C

EhaG Hep-Hag: a.g.n(s/a)iA(i/l)G..s.a.g.ns(v/i)A(l/i)Gn.s.

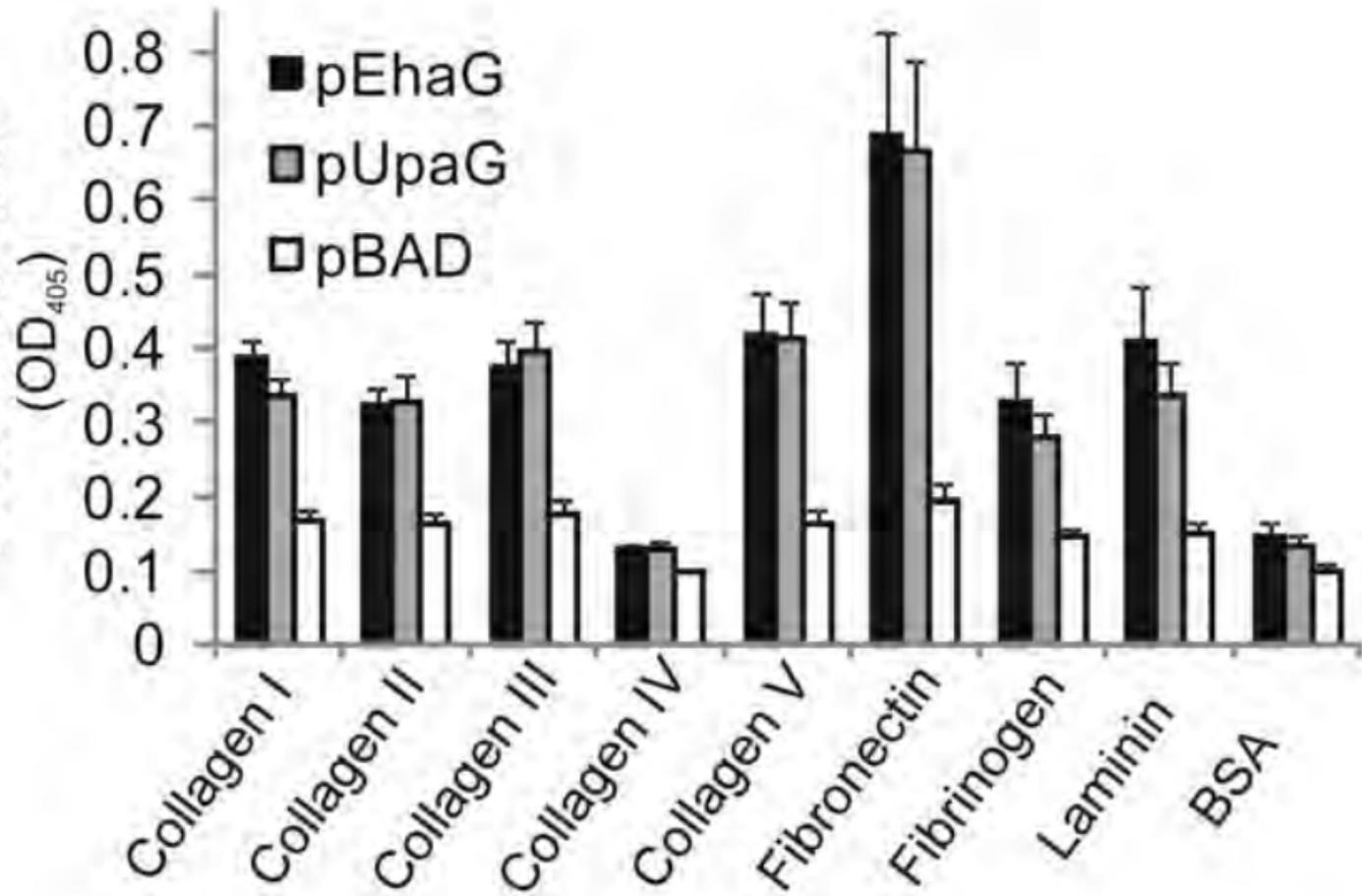
UpaG Hep-Hag: a.g..s(i/l)A(l/i)G.(n/k)(s/a).a.g..(s/a)(v/i)A(l/i)G..s.

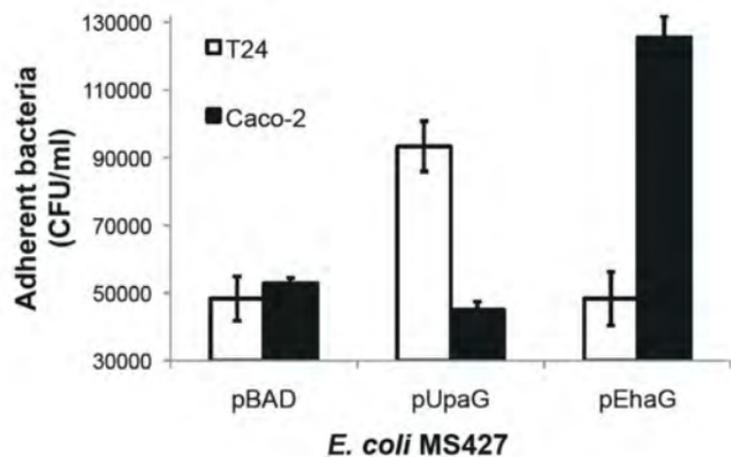
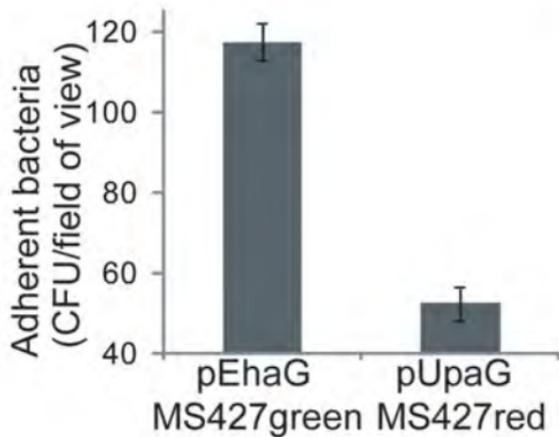
EhaG Him: s(k/v)ITNVa(d/n)G.(i/v)(s/t)..s(s/t)DA(v/i)NgsQL

UpaG Him: s(k/v)ITNVa(d/n)G.(i/v)(s/t)..s(s/t)DA(v/i)NgsQL

A**B****C**

Bacterial Adhesion



A**B****C**