Microbiology (2016), **162**, 1091–1102 **DOI** 10.1099/mic.0.000303

Self-association motifs in the enteroaggregative *Escherichia coli* heat-resistant agglutinin 1

Jessica Glaubman, Jennifer Hofmann, Megan E. Bonney, Sumin Park, Jessica M. Thomas, Bashkim Kokona, Laura I. Ramos Falcón, Yoonjie K. Chung, Robert Fairman and Iruka N. Okeke†

Department of Biology, Haverford College, Haverford, PA, USA

Correspondence
Iruka N. Okeke
iruka.n.okeke@gmail.com

The heat-resistant agglutinin 1 (Hra1) is an integral outer membrane protein found in strains of *Escherichia coli* that are exceptional colonizers. Hra1 from enteroaggregative *E. coli* strain 042 is sufficient to confer adherence to human epithelial cells and to cause bacterial autoaggregation. Hra1 is closely related to the Tia invasin, which also confers adherence, but not autoaggregation. Here, we have demonstrated that Hra1 mediates autoaggregation by self-association and we hypothesize that at least some surface-exposed amino acid sequences that are present in Hra1, but absent in Tia, represent autoaggregation motifs. We inserted FLAG tags along the length of Hra1 and used immune-dot blots to verify that four *in silico*-predicted outer loops were indeed surface exposed. In Hra1 we swapped nine candidate motifs in three of these loops, ranging from one to ten amino acids in length, to the corresponding sequences in Tia. Three of the motifs were required for Hra1-mediated autoaggregation. The database was searched for other surface proteins containing these motifs; the GGXWRDDXK motif was also present in a surface-exposed region of Rck, a *Salmonella enterica* serotype Typhimurium complement resistance protein. Cloning and site-specific mutagenesis demonstrated that Rck can confer weak, GGXWRDDXK-dependent autoaggregation by self-association. Hra1 and Rck appear to form heterologous associations and GGXWRDDXK is required on both molecules for Hra1–Rck association. However, a GGYWRDDLKE peptide was not sufficient to interfere with Hra1-mediated autoaggregation. In the present study, three autoaggregation motifs in an integral outer membrane protein have been identified and it was demonstrated that at least one of them works in the context of a different cell surface.

Received 10 December 2015
Accepted 5 May 2016

INTRODUCTION

Bacteria that colonize mucosal surfaces can typically adhere to eukaryotic cells and other bacteria. Adhesive phenotypes in *Escherichia coli* and related bacteria are conferred by surface structures known as pili, by secreted proteins and by non-structural outer membrane proteins. The latter include self-associating autotransporters (SAATs) (Klemm *et al.*, 2006), as well as integral membrane proteins that are often homologous to the major surface proteins of *E. coli*. Hra1 was originally described as an autoaggregating and hemagglutinating protein from an O9 : H10 : K99 *E. coli* strain pathogenic in piglets and lambs (Lutwyche *et al.*, 1994). This protein and the highly similar Hek have since been found in *E. coli,* which cause invasive infections and diarrhoea in humans (Cooke *et al.*, 2010; Mancini *et al.*, 2011; Srinivasan *et al.*, 2003). While Hra1 is also found in non-pathogenic *E. coli*, its presence enhances colonization of the diarrhoeagenic *E. coli* category enteroaggregative *E. coli* (EAEC) and of invasive *E. coli* (Bhargava *et al.*, 2009; Fagan & Smith, 2007; Mancini *et al.*, 2011). Hra1 is an integral outer membrane β-barrel protein. It is predicted to consist of eight β -strands containing alternating hydrophobic residues, flanked by aromatic residues. The membrane-embedded β-strands are believed to be interspersed by seven loops. Topological modeling by Fagan *et al.* (2008) predicted that the four longest, unstructured and charged loops are surface exposed, while the other three loops form short periplasmic turns. These structural features are common to the major integral outer membrane proteins in *E. coli* and related bacteria (Bhargava *et al.*, 2009). Hra1 is one

†Present address: Department of Pharmaceutical Microbiology, Faculty of Pharmacy, University of Ibadan, Ibadan, Nigeria

Abbreviations: EAEC, enteroaggregative *Escherichia coli*; SAAT, self-associating autotransporter; Hra1, Heat resistant agglutinin 1; TFA, trifluoroacetic acid; Hra2, Heat resistant agglutinin 2.

One supplementary table and three supplementary figures are available with the online Supplementary Material.

of a family of OmpA-like 'agglutinin' proteins, which includes the invasin Tia and the adhesin Hra2 (Fleckenstein *et al.*, 1996; Mancini *et al.*, 2011). Agglutinin proteins are highly conserved, sharing over 65 % identity overall, but vary considerably within their predicted surface-exposed loops (Mancini *et al.*, 2011). Hra1, Hra2 and Tia all confer adherence to eukaryotic cell surfaces and Hra1 and Tia have been shown to bind to heparinated proteoglycans (Bhargava *et al.*, 2009; Fagan & Smith, 2007; Fagan *et al.*, 2008; Mammarappallil & Elsinghorst, 2000; Mancini *et al.*, 2011). Proteoglycan binding has been mapped independently to regions of both proteins that overlap within a conserved portion of surface-exposed loop 2 (Fagan *et al.*, 2008; Fleckenstein *et al.*, 2002). Previous experiments in our laboratory and in the laboratories of others demonstrate that Hra1, Hra2 and Tia are sufficient to confer adherence on a laboratory strain of the species *E. coli*. The agglutinin genes are also sufficient to confer other phenotypes, such as invasion for Tia and autoaggregation for Hra1 (Bhargava *et al.*, 2009; Fagan & Smith, 2007; Fagan *et al.*, 2008; Fleckenstein *et al.*, 2002; Lutwyche *et al.*, 1994; Mancini *et al.*, 2011).

Inter-bacterial adherence or autoaggregation could promote colonization by allowing non-adherent bacteria to hitchhike onto adherent ones. It may also confer physical protection from adversity *in vitro* and *in vivo* and increase the effective infecting dose per particle. The relevance of autoaggregation to disease or colonization has been demonstrated experimentally for some pathogens, but even where it lacks physiological relevance, autoaggregation may enhance survival in a harsh environment.

Hra1 was originally identified due to its autoaggregation phenotype (Lutwyche *et al.*, 1994), but this has been less well studied than the eukaryotic cell interactions it mediates. Moreover, although a number of unrelated outer surface proteins elicit autoaggregation, ranging from cell-wall anchored proteins in Gram-positive organisms (Frick *et al.*, 2000) to Gram-negative SAATs, very few specific motifs required for this phenotype have been defined. Among Gram-negative integral membrane proteins, very little is known about the residues contributing to, or the mechanisms of, autoaggregation. Fagan *et al.* (2008) systematically deleted the four surface-exposed loops of Hra1, replacing them with alanine-arginine bridges. They were able to show that the loop 1 mutant was unimpaired with respect to autoaggregation, but that deletion of loops 2, 3 and 4 reduced autoaggregation to varying degrees. The expression and surface exposure of Hra1 mutants containing the bridges, particularly the loop 3 deletion, was however reduced compared to wildtype, which may have contributed in part to the reduced autoaggregation (Fagan *et al.*, 2008). Additionally, as whole loops were deleted, the contributions of specific motifs within those loops is unknown.

In a recent report we demonstrated that Tia does not confer autoaggregation to the extent of Hra1 (Mancini *et al.*, 2011). In the present study, we have verified that predicted surface loops of Hra1 are indeed surface exposed. By then swapping specific surface-exposed motifs in Hra1 loops 2, 3 and 4 with their cognate motifs in Tia, we can hypothesize that at least some of these motifs confer autoaggregation.

METHODS

Strains and plasmids. The *hra1* and *tia* genes were previously cloned from EAEC strain 042 and enterotoxigenic *E. coli* strain H10407, respectively (Bhargava *et al.*, 2009; Mancini *et al.*, 2011) (Table 1). Identical *rck* alleles were cloned in this study from genomic DNA obtained from *Salmonella enterica* serotype Typhimurium strains SF1005 [an attenuated, *rpoS*-negative variant of strain SL1344 (Fang *et al.*, 1992)] and NCTC 5710. *E. coli* TOP10 or DH5α (Life Technologies) were used as hosts for constructed plasmids. The plasmids used or constructed for this work are listed in Table 2. Unless otherwise indicated, strains were cultured in Luria Broth (LB) or LB agar at 37 °C and maintained at −80 °C in LB : glycerol 1 : 1; ampicillin was added at a concentration of 100 μg ml to maintain plasmid selection.

Cloning and general molecular biology procedures. Standard molecular biological methods were employed in the present study (Sambrook & Russell, 2001). PCR reactions used 1 μM oligonucleotide primer in each reaction and were templated with genomic DNA extracted using an Easy DNA kit (Life Technologies). All amplifications began with a two-minute hot start at 94 °C followed by 30 cycles of denaturing at 94 °C for 30 s, annealing for 60 s at 10 °C (below

Table 1. Sources of strains for the agglutinin genes investigated in this study

EAEC, enteroaggregative *E. coli*.

Source strain	Agglutinin gene	Pathotype and source	References/Source
H10407	*tia*	Enterotoxigenic *E. coli* isolate from Bangladesh	Evans *et al.*, 1975; Fleckenstein *et al.*, 1996; Mancini *et al.*, 2011
042	*hra1*	EAEC isolate from Peru	Bhargava *et al.*, 2009; Chaudhuri *et al.*, 2010; Nataro *et al.*, 1985
60A	*hra2*	EAEC isolate from Mexico	Czeczulin *et al.*, 1999; Mancini *et al.*, 2011
SF1005	*rck*	Attenuated, *rpoS*-negative variant of strain *Salmonella enterica* serotype Typhimurium SL1344	Fang *et al.*, 1992
NCTC 5710	*rck*	*S.* Typhimurium	National Collection of Type Cultures

Table 2. Plasmid vectors, transposon and primary clones used in this study

DNA molecule	Description	Source
EZ-Tn5<Not I/KAN-3>	Tn*5* for *in vitro* transposition containing kanamycin resistance gene flanked by *Not*I sites	Epicentre
pBAD/Thio	Arabinose expression vector. AmpR	Life Technologies
pBADLacZ	*lacZ* in pBAD/Thio	Life Technologies
pBR322	Ampicillin and tetracycline-resistant general cloning vector	New England Biolabs
pBJ1	*hra1* from enteroaggregative *E. coli* strain 042 cloned into pBR322	Bhargava *et al.*, 2009
pJDHra1	*hra1* from enteroaggregative *E. coli* strain 042 cloned into pBAD/Thio	Mancini *et al.*, 2011
pJDTia	*tia* from enterotoxigenic *E. coli* strain H10407 cloned into pBAD/Thio	Mancini *et al.*, 2011
pSP5710	*rck* from *Salmonella typhimurium* strain NCTC5710 cloned into pBAD/Thio	Present study
pGEX-KT :: myoII-FLAG	N-terminal FLAG-tagged myosin II from *Acanthamoeba castellanii* cloned into the *Bam*HI site of the expression plasmid pGEX-KT.	Zolkiewski *et al.*, 1997

the primer annealing temperature) and extending at 68 °C for 2-min for every kilobase of DNA. The *rck* gene was amplified for expression in the pBAD/Thio vector (Life Technologies) using purpose-designed primers 5′-AAAAAAATCGTTCTGTCCTCACT-3′ and 5′-GAAAAGGTAACCGACACCAA-3′and *Pfx* polymerase (Life Technologies). All clones and mutants were sequence-verified. Following two hours of induction with 2 % arabinose (Sigma), or repression with 2 % glucose (Sigma), expression and outer membrane localization of all agglutinin and *rck* clones in *E. coli* TOP10 were verified by SDS-PAGE analysis of the outer membrane fractions of harvested bacteria purified by the method of Chart *et al.* (1997).

Insertion of FLAG tags along the length of Hra1. Plasmid pBJ1 was incubated with transposon with EZ-Tn5 <Not I/KAN-3> and transposase, using an EZ-Tn5 in-frame linker kit (Epicentre), according to the manufacturer's instructions. The resultant transposon mutagenesis library was transformed into *E. coli* DH5α and plated onto LB-containing neomycin. Restriction analysis with *Sph*I and *Ssp*I identified 49 clones with insertions within the *hra1* open-reading frame. Kanamycin/neomycin resistance cassettes were excised from these clones by *Not*I digestion and replaced with synthesized DNA encoding the FLAG sequence, DYKDDDDK. The FLAG tag with *Not*I-compatible overhangs was prepared by annealing the oligonucleotides FLAGlink1 (5′-GGCCGCGACTACAAGGACGATGACGATAAG-3′) and FLAGlink2 (5′-GGCCCTTATCGTCATCGTCCTTGTAGTCGC-3′), and phosphorylating the double-stranded product prior to ligation. Upon sequencing, 16 clones with in-frame insertions across the whole length of the gene with $\leq$28 residues between clones were identified. These plasmids were transformed into DH5α and the outer membrane localizations of the FLAG-tagged *hra1* gene products were verified in all but two of the clones by Western blotting with anti-DYKDDDDK tag antibodies (Cell Signaling). Surface localization of FLAG tags was determined by dot blotting using the same antibodies. Each clone was tested on three or more different occasions. DH5α carrying pBR322 and a cloned version of myosin II (Zolkiewski *et al.*, 1997), which had been N-terminal FLAG-tagged, were used as positive and negative controls, respectively, for FLAG detection.

Sequence analyses. Pairwise sequence comparisons were made using BLAST, applying default parameters for short-sequence analysis where appropriate (Altschul *et al.*, 1990). Multiple sequence alignments were performed using CLUSTALW (Chenna *et al.*, 2003) in MEGA version 6 (Tamura *et al.*, 2013). Neighbour-joining trees were reconstructed from CLUSTAL alignments also using MEGA and bootstrapped after 500 samplings. Topological modeling was performed using HHpred (Soding *et al.*, 2005) and PredictProtein (Yachdav *et al.*, 2014) and depicted using TOPO2 (http://www.sacs.ucsf.edu/TOPO2/).

Site-directed mutagenesis. Mutations were made in cloned genes using a QuickChange mutagenesis kit (Stratagene) in accordance with manufacturer's instructions, with the exception that the *Pfu* polymerase provided with the kit was substituted with *Pfx* polymerase (Life Technologies). Primers were designed as recommended by the QuickChange manufacturer with four or fewer nucleotides changed. For larger substitutions we employed the strategy of Deng *et al.* (2007), increasing the length of the sequence flanking the mutation to 10–20 bp, the primer concentrations to 10 ng μl^{-1} and the template concentrations to 1–2 ng μl^{-1}. Alterations made and primer sequences are presented in Table S1 (available in the online Supplementary Material). All mutant clones were verified by double-stranded Sanger sequencing.

Autoaggregation assay. Autoaggregation was quantified as bacterial settling in overnight cultures of LB over time as described by Hasman *et al.* (1999). All plasmids were tested in an *E. coli* TOP10 background and assays were performed in duplicate on at least three separate occasions with each strain tested by at least two researchers. Cultures of each strain were adjusted to the same optical density at 600 nm (OD_{600}). For each assay, 8 ml of each adjusted culture was placed into two separate tubes. One tube remained static and the other was vortexed before each optical density measurement. The tubes were incubated without shaking at 37 °C. At designated time points, 0.5 ml was removed from within 2 cm of the surface of the culture and the OD_{600} of the sample measured. Statistical comparisons between OD_{600} supernatant measurements for different strains was performed using a one- or two-tailed Student's *t*-test as appropriate.

Microscopy observation of autoaggregated bacteria. Statically incubated LB cultures in which autoaggregation had occurred overnight were sampled just above the pellet at the base of the tube. A 20 µl sample of this culture was placed into a 20 µl drop of phosphate buffered saline on a slide. The suspension was air-dried and then heat-fixed by passing through a Bunsen flame. The resulting fixed smear was stained with Gram's safranin and observed by light microscopy under oil immersion (1000x) using a Nikon Eclipse 80i light microscope.

Peptide synthesis and spectral analyses Hra1-derived peptide GGYWRDDLKE was made on an Applied Biosystem 433A peptide synthesizer utilizing standard Fmoc chemistry and using PAL resin (Advance ChemTech), which provides an amide group at the carboxy-terminus upon cleavage from the support. The peptide was acetylated at the amino terminus prior to cleavage off the resin with trifluoroacetic acid (TFA). The crude peptide was dissolved in water and HPLC-purified using a Varian ProStar system equipped with a Varian Dynamax semi-preparative C18 column. Linear gradients were run using water as the polar solvent and acetonitrile as the nonpolar solvent, each

containing 0.1% TFA. The peptide eluted in 31% acetonitrile. The identity of the purified peptide was confirmed by MALDI-TOF mass spectrometry, yielding a molecular mass of 1280.8 Da. The theoretical molecular weight of GGYWRDDLKE is 1280.4 Da, with an overall charge of −1 at neutral pH and an isoelectric point of 4.3. The extinction coefficient at 278 nm is 6970 M^{-1} cm^{-1}. The circular dichroism spectra of the peptide at 1 mM was determined in 2.7 mm KCl, 136.9 mM NaCl, 1.5 mm KH_2PO_4, and 89 mM Na_2HPO_4 pH 7.1, using an Aviv 410 circular dichroism spectropolarimeter (Aviv Biomedical). Infrared spectra were determined using a Vertex 70 instrument (Bruker Optics) equipped with Opus version 5.5 software and using a 10 mm sample after 24 h of incubation. Analytical ultracentrifugation sedimentation velocity experiments were performed using a Beckman model Optima XL-A AUC instrument (Beckman Coulter) equipped with an An-60 Ti rotor. The sedimentation boundary was measured at a speed of 50 000 rpm at 20 °C using a step size of 0.003 cm, a delay time of 0 s and a total of 60 scans. Model-independent continuous c(s) distribution analysis to determine the heterogeneity of Hra-1-derived peptide was performed using Sedfit v.14.4d. Data, fits, residues and c(s) distributions were plotted using Gussi interface v.1.0.3 implemented in Sedfit v.14.4d (Dam *et al.*, 2005).

RESULTS

Hra1 mediates autoaggregation by self-association and does not associate with its close homologue Tia

The present and other authors have reported that Hra1 (and Hek) are sufficient to confer autoaggregation on *E. coli* (Bhargava *et al.*, 2009; Fagan *et al.*, 2008; Mancini *et al.*, 2011). This autoaggregation occurs at a slower rate compared to that produced by SAATs [visible within 2 h (Hasman *et al.*, 1999; Sherlock *et al.*, 2004)], but is clearly visible after cultures have stood for several hours. Interestingly, although Hra1 is 67 % identical to Tia, Tia confers comparably weak autoaggregation, which is quantitatively insignificantly different to that seen in strains of *E. coli* carrying the vector alone (Fig. 1) (Mancini *et al.*, 2011). We hypothesized that like some other autoaggregants, such as the SAATs antigen 43, AIDA-1 and Tib, as well as Protein H of *Streptococcus pyogenes*, and the *Bordetella pertussis* filamentous hemagglutinin (Frick *et al.*, 2000; Klemm *et al.*, 2006; Menozzi *et al.*, 1994), Hra1 might mediate autoaggregation by self-association. To test this hypothesis, in the present study the autoaggregation has been compared between cultures containing *E. coli* expressing Hra1 alone, or *E. coli* expressing Hra1 mixed with an equal proportion of *E. coli* not expressing Hra1, but instead making LacZ from the same vector. As shown in Fig. 1a, the mixed cultures autoaggregated to an extent intermediate between the cultures expressing Hra1 alone or LacZ alone, suggesting that Hra1-expressing cells autoaggregate with one another and not with *hra1*-negative bacteria. We hypothesized that Tia, which is very similar to Hra1, might retain the ability to interact with Hra1, even though this protein is not itself sufficient to confer autoaggregation. When we tested Hra1:Tia-expressing mixtures, however, we found that they autoaggregated to the same extent as Hra1:LacZ expressing mixtures (Fig. 1). Therefore, Hra1 is a self-associating autoagglutinin that does not associate with its close homologue Tia.

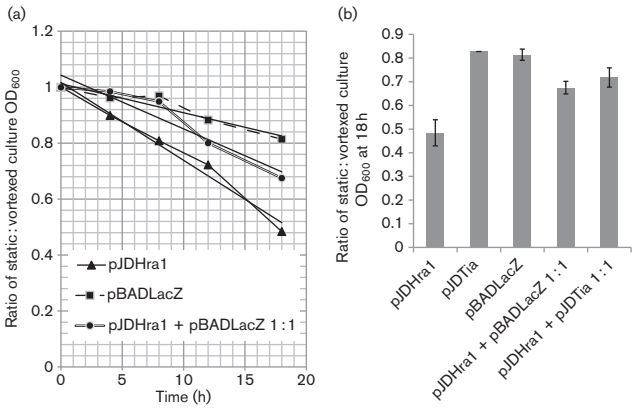

Fig. 1. Hra1 mediates autoaggregation by self-association. Autoaggregation was quantified as the ratio of absorbance of bacteria at the top of the culture in paired static and vortexed tubes. (a) Mixed cultures of bacteria (1 : 1) expressing Hra1 and LacZ from the arabinose promoter autoaggregate half as much as cultures containing bacteria expressing Hra1 alone. Superimposed on the plots are linear trend lines to enable comparisons. Autoaggregation-mediated declines in the absorbance of static cultures of the strain carrying the Hra1 clone alone were significantly lower than those from the 1 : 1 mixed culture after 8, 12 and 18 h ($P<0.05$). (b) Mixed cultures (1 : 1) of bacteria expressing Hra1 and Tia demonstrate that Hra1 does not mediate association with its close homologue Tia. Differences between Hra1 and all the other samples were significant ($P<0.05$). Measurements of Tia were not significantly different from those of LacZ or the Hra1 : Tia mix.

Inserted FLAG tags verify the predicted topological model for Hra1 *in silico*

Fagan *et al.* (2008) predicted that Hra1 is an integral outer membrane protein with four surface-exposed loops and three periplasmic turns. This model is similar to the one predicted earlier for Tia (Mammarappallil & Elsinghorst, 2000) and analogous to those of OmpA and related proteins for which structural information is available. We verified the model using HHpred (Soding *et al.*, 2005) and PredictProtein (Yachdav *et al.*, 2014) (Fig. 2). Overall, the model presents a similar picture to those from earlier investigators, with the important exception that the outward-leading β-sheet is much longer. In reality, the longer β-strand is unlikely to be lodged within the membrane and it is probable that the second periplasmic turn and/or the third external loop are longer than the *in silico* models predict. We performed dot blots on varieties of Hra1 with FLAG tags inserted at various locations along the length of the protein. Six of the insertions were predicted to be located in loops (at least one in each loop) and the rest in membrane-spanning regions or periplasmic turns. The

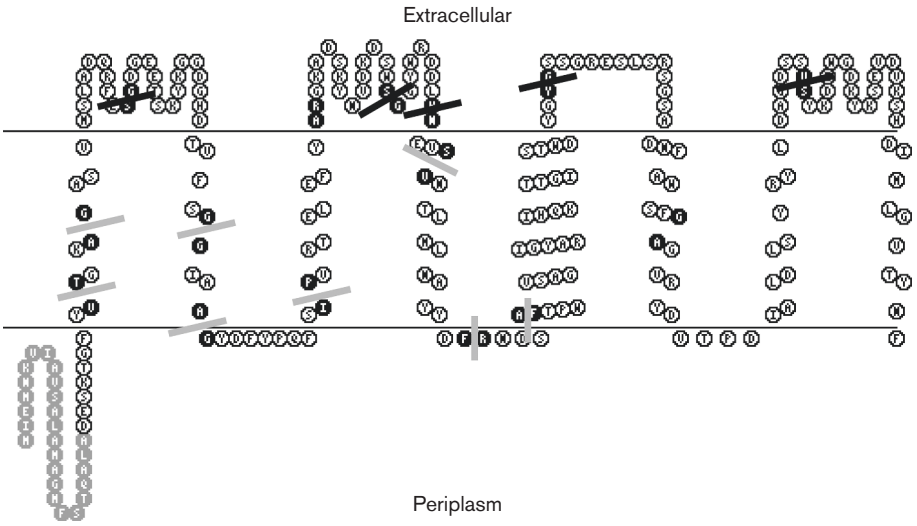

Fig. 2. Topological model of Hra1 showing the four predicted extracellular loops and three predicted periplasmic turns. The N-terminal signal sequence is marked in grey. In this study FLAG tags were inserted between 16 pairs of adjacent residues marked in black. Bars between the residues indicate those insertions for which tagged protein could be detected by anti-FLAG antibody. Sites marked with grey bars could be detected by SDS-PAGE followed by Western blotting alone, whilst those with black bars could be detected by dot blotting of unlysed cells as well as by Western blotting of outer membrane preparations.

first insertion, nine residues downstream from the signal sequence cleavage site verified experimentally by Fagan *et al.* (2008), and therefore too proximal to the N-terminus of the protein to transverse the membrane, served as a non-surface-exposed control. All the FLAG-tagged versions were expressed, but none showed autoaggregation significantly above baseline levels (data not shown), suggesting that even insertions as short as 30 nucleotides functionally disrupt Hra1-mediated autoaggregation. As shown in Fig. 2 all six of the loop-located tags, but none of the remaining ten, could be surface-detected with antibody. So, data support the topological prediction and suggest that residues contributing to autoaggregation are likely to lie within the predicted surface-exposed loops.

Specific surface-exposed residues in Hra1are required for autoaggregation

Hra1 and Tia are highly similar proteins that are functionally different in their autoaggregation phenotypes, providing us with the opportunity to use their sequences as the basis for identifying autoaggregation-specific residues and motifs. Fagan *et al.* (2008) previously showed that the Hra1 homologue Hek has four predicted surface-exposed loops and that loops 2 and 4 are required for full autoaggregation. Their study demonstrated that loop 1 is not required for autoaggregation and the other loops likely contribute to this phenotype, but as some of their deletion mutants were not expressed optimally the contributions from the other three loops and specific motifs within them, to this phenotype, remain open to question.

Motif-finding studies that use substitutions, which classically replace residues of interest with 'neutral' amino acids (such as alanine), are typically less drastic than deletion analyses. Little is known about the structure–function of autoaggregation and rather than make presumptions about neutrality, we elected to swap out residues or groups of residues unique to Hra1 in loops 2–4 with the cognate amino acids present in Tia (Fig. 3). We used a combination of point mutations and domain-swapping (Deng *et al.*, 2007) to effect the changes specified in Fig. 3. and Table S1, and verified them by sequencing. SDS-PAGE of Coomassie-stained outer membrane preparations verified that equivalent amounts of Hra1, Tia and the Hra1 mutants, carrying different Tia motifs, were present in the outer membranes of test strains (data not shown). As shown in Fig. 4, swapping out four of the residue/motif sets investigated did not alter the ability of Hra1 to autoaggregate. However, three sets of sequence changes resulted in autoaggregation, which was significantly impaired when compared to Hra1.

Fagan *et al.* found that loop 2 was required for autoaggregation, adherence and invasion (Fagan *et al.*, 2008). We identified two non-contiguous motifs on this loop required for autoaggregation based on the Fagan *et al.* topological model (Fagan *et al.*, 2008), and that in Fig. 2 these motifs would represent the outward and inward sections of that loop. Although the intervening region between the motifs was not at all conserved, swapping the NVDK-DSWS (residues 112–121) in Hra1 for TLDTWRSPM in Tia had no discernible effect on autoaggregation. Thus, these residues may act as spacers or contribute to other Hra1 phenotypes. In

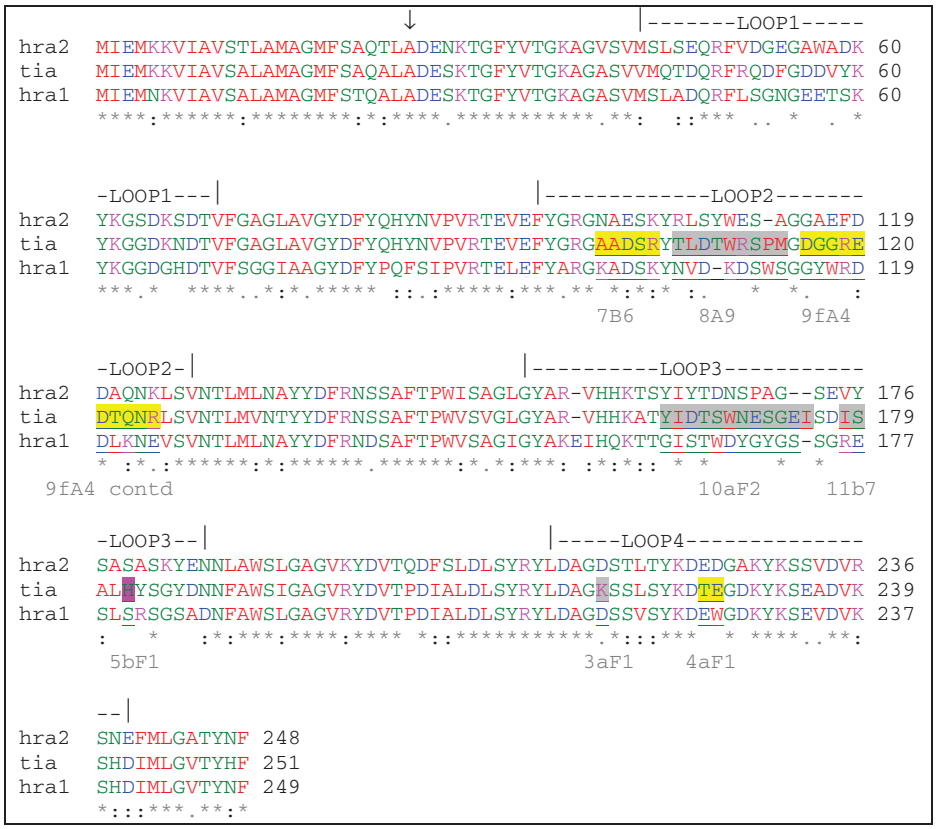

Fig. 3. Multiple alignment of Hra1, Tia and Hra 2 indicating the four variable putative surface-exposed loops. An arrow marks the probable signal sequence cleavage site. Underlined residues in the Tia sequence were used to replace the corresponding underlined portions of Hra1. Tia motifs that produced reduced autoaggregation in the Hra1 background are highlighted in yellow and the S178H substitution that increased autoaggregation is highlighted in pink. Substitutions that produced no phenotype when Hra1 residues were replaced with the corresponding Tia motif are highlighted in grey. Designation of specific pJG series clones for each mutant are listed below the alignment in grey.

contrast, swapping residues 98–102 (KADSK) on Hra1 to AADSR on the outward-leading portion of the loop, as occurs in Tia, significantly reduced autoaggregation ($P<0.0001$). The most obvious explanation for this finding is that reducing the proportion of basic residues in this motif may have interfered with electrostatic interactions that could contribute to autoaggregation. Indeed Hra1 has a low pI: 4.78, which is slightly lower than that of Tia (pI=5.21), and produces a negative charge on the cell surface at physiological pH (Lutwyche *et al.*, 1995). Lutwyche *et al.* (1994), who first described Hra1 from an animal pathogen noted that its net negative charge suggested that autoaggregation was unlikely to be mediated by electrostatic charges. It is unlikely that electrostatic effects can solely account for Hra1-mediated autoaggregation, but the importance of the KADSK motif suggests that they have a contributory role.

If electrostatic interactions do contribute to Hra1 autoaggregation, autoaggregation dynamics should vary with ionic strength or the salt concentration of the surrounding environment (Klemm *et al.*, 2004). We measured the autoaggregation of Hra1 in the presence and absence of sodium chloride. We found that bacteria expressing Hra1 autoaggregate in the presence of 150 mm sodium chloride, but fail to do so in salt-free LB. Furthermore, autoaggregation is greatly reduced in 50 mm sodium chloride (Fig. S1). This finding, along with the requirement for the basic extracellular KADSK motif in a largely acidic protein, suggests that electrostatic interactions may affect autoaggregation. However, as Hra1 does not associate with Tia, it is likely that other interactions also contribute to this phenotype. Of particular note is that swapping the motif on the inward leading part of loop2 (as discussed in more detail later on) did not appreciably alter the predicted pI.

A Glu-Trp to Thr-Gln change in loop 4 (residues 226–227 in Hra1) also severely impaired autoaggregation. Unexpectedly, a loop 3 Ser178His alteration in Hra1 produced a mutant that autoaggregated more than the wildtype. The presence of a histidine residue at the equivalent position on loop 3 in Tia, in the absence of the required motifs in loops

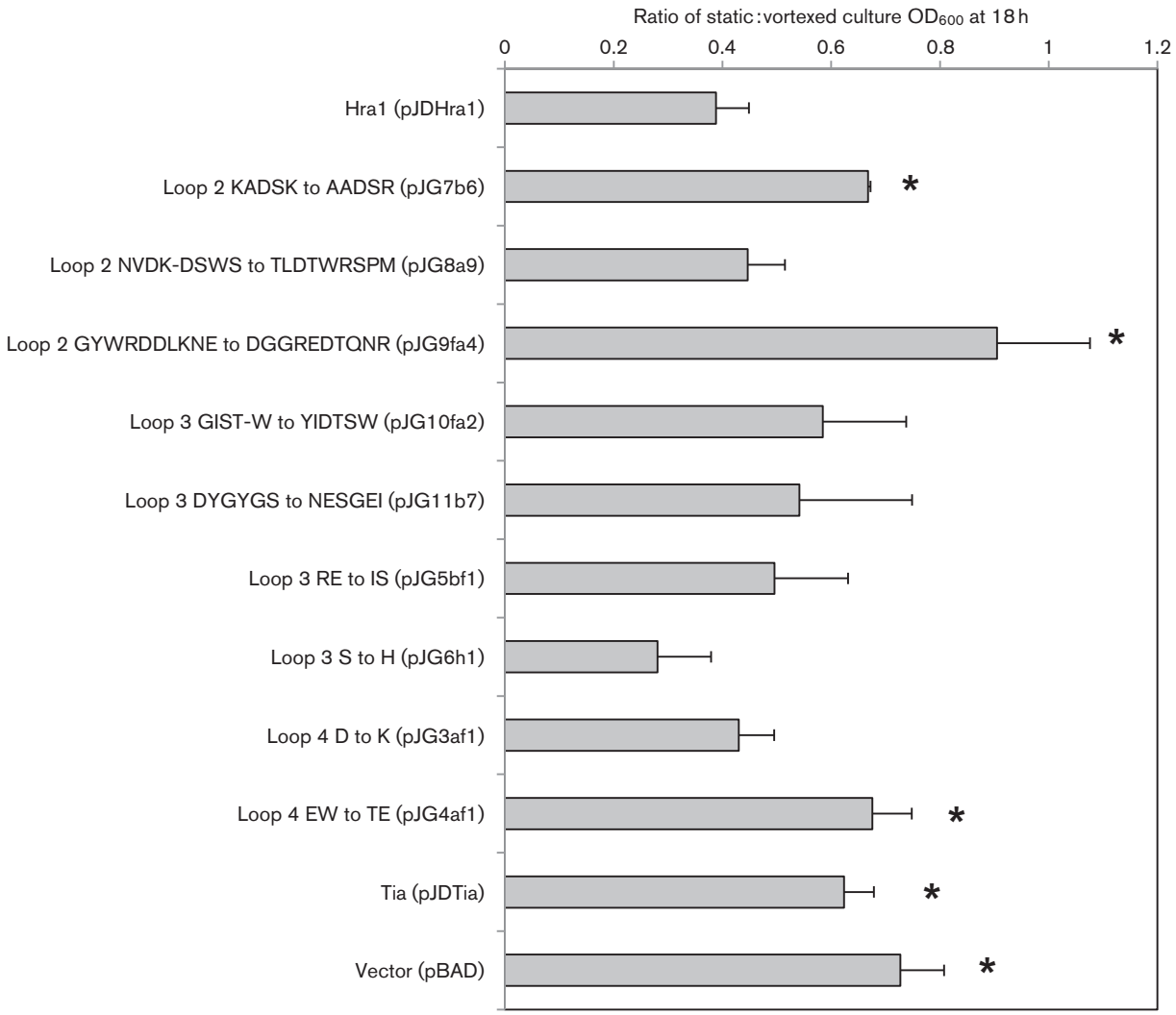

Fig. 4. Autoaggregation of cultures expressing Hra1, Tia and Hra1 derivatives with nine different swapped-in Tia domains. Autoaggregation is quantified as the ratio of absorbance of bacteria at the top of the culture in paired static and vortexed tubes at 18 h. Asterixed bars were significantly greater than the values obtained with the Hra1 clone pJDHra1 (one- or two-tailed Student's *t*-test, as appropriate).

2 and 4, may account for the weak autoaggregation seen with this in the protein when expressed from higher copy-number clones.

An Hra1 autoaggregation motif is found in the *Salmonella* Typhimurium plasmid-encoded outer membrane protein Rck

No function was associated with the potential autoaggregation motifs detected within this study in the PROSITE database. We performed short-sequence protein BLASTs with 6–8 residue sequences containing autoaggregation motifs that we had identified in Hra1, including flanking sequences and/or unnamed residues (Xs) where necessary. We expectedly recovered Hra1 and Hek alleles in this search as well as a number of hypothetical proteins, and even weak eukaryotic hits, which is not surprising given the short query sequences. We examined our results list for putative surface-exposed bacterial proteins. These included a bacteriophage protein of unknown function from *Oxalobacter formigenes* (EEO27164) and Rck, an outer surface invasin that has been previously reported from *Salmonella enterica* subspecies Typhimurium. Rck (resistance to complement killing), is encoded on the virulence plasmid and confers serum resistance (Heffernan *et al.*, 1992a, b). Like the agglutinins, it is predicted to be composed of four surface-exposed loops interspersed with membrane inserted β-strands. The serum resistance phenotype has been mapped to the first and third of Rck's four surface-exposed loops.

Rck homologues include the bacteriophage lambda-encoded *E. coli* outer membrane protein Lom (Barondess & Beckwith, 1990), OmpX (including *Yersinia* alleles of this gene, referred to as Ail), as well as chromosomally encoded *Salmonella enterica* subspecies Typhimurium outer membrane protein PagC (Heffernan *et al.*, 1992a). As shown by the dendogram in Fig. 5a, Rck is much less similar to Hra1 and Tia. Outer surface proteins belonging to the Rck/PagC/Ail family are 17–19 KDa in size, as opposed to the agglutinin family that, after processing, are about 25 KDa. Rck is the only member of its family containing the GGXWRDDXK motif. Of those that have been characterized functionally, *rck* confers serum resistance and invasion, *pagC* is necessary for intracellular survival and *ompX/ail* confers a number of phenotypes including adherence, invasion, serum resistance and autoaggregation (Kolodziejek *et al.*, 2010). So, this is another integral outer-membrane protein family composed of members that confer different colonization-associated phenotypes.

We screened *S. typhimurium* strains SF1005 and NCTC 5710 for the presence of *rck* by PCR, and for autoaggregation. The two strains of species of the genus *Salmonella* did not autoaggregate under the test conditions, but did harbour the *rck* gene. Since Hra1-mediated autoaggregation cannot be measured *in vitro* in wildtype EAEC strain 042, most likely due to steric masking from other surface factors (Bhargava *et al.*, 2009), we elected to determine whether Rck was sufficient to confer autoaggregation in a heterologous background.

We cloned and expressed the *rck* gene from *S. typhimurium* strain NCTC 5710 under the control of the arabinose promoter. The predicted amino acid sequence of the allele was identical to that of strain SF1005 and to Rck from S. *typhimurium* strains SL1344 and LT2 in the Genbank database (Accession numbers YP_006203728 and NP_490501, respectively). When expressed in an *E. coli* strain from the arabinose promoter, Rck was sufficient to confer autoaggregation, but at a slower rate than Hra1 (Fig. 5; $P<0.05$). An aliquot of bacteria sampled from the top of the pellet of an overnight static liquid culture in which pBAD/Thio clones had been induced with arabinose was fixed onto a slide and stained with safranin. When viewed at a magnification of ×1000, bacteria sampled from pBADLacZ overnight cultures existed as single bacteria or occasionally in aggregates of 2–6 bacteria. As shown in Fig. 5c from pJDHra1 (Hra1-expressing) most bacteria drawn from just above the pellet of autoaggregated cultures were part of comparatively massive bacterial aggregates. When we searched in regions adjacent to these aggregates, we found that the majority of bacteria outside large aggregates were in smaller dense aggregates of several dozen to a few hundred cells; there were very few single cells. By contrast, bacteria from pSP5710c (Rck-expressing) cultures were predominantly in small aggregates with considerably more single bacteria also visible. The aggregates were smaller in size and less dense than Hra1-mediated aggregates and this may account for their slower settling rate

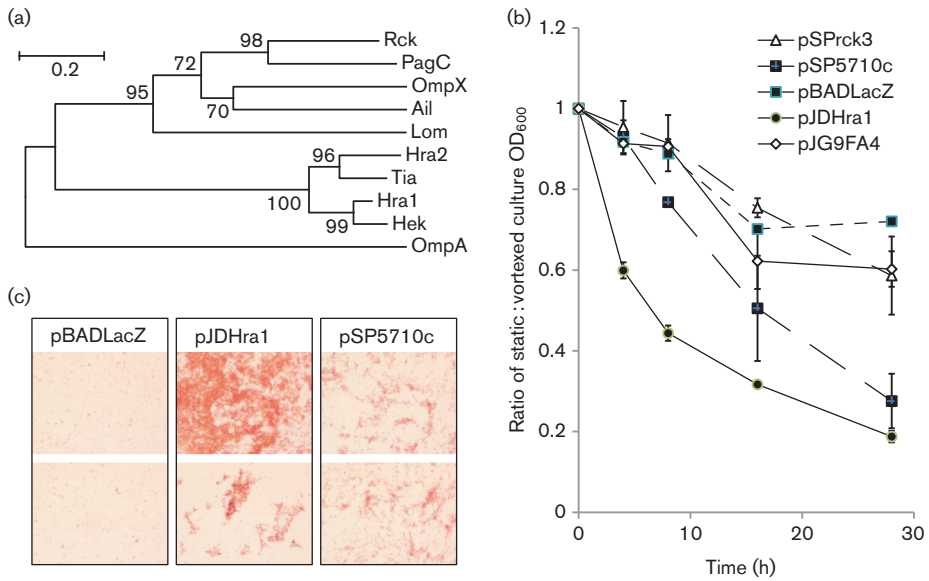

Fig. 5. Rck is sufficient to mediate autoaggregation. (a) Neighbour-joining tree reconstructed from a multiple alignment of amino acid sequences of the agglutinin family and the Rck family of integral outer membrane proteins. Bootstrap values are indicated at the tree nodes. (b) Autoaggregation mediated by Hra1 (from pJDHra1) and Rck (from pSP5710c) and their GGXWRDDXK-mutated derivatives (pJG9FA4 and pSPrckA3 respectively). A clone expressing LacZ from the same vector (pBADLacZ) was employed as a negative control. (c) Bacteria from settled cultures of *E. coli* expressing *LacZ*, Hra1 and Rck as visualized by light microscopy after staining with safranin. Two representative fields are shown for each culture. The scale-bar indicates distance in amino acid substitutions per position.

(Fig. 5c). We mutated the GGXWRDDXK motif in Rck, replacing it with Tia's cognate sequence and found that the mutated allele (in pSPrck3), like the Hra1 GGXWRDDXK mutant (pJG9FA4) failed to confer autoaggregation (Fig. 5; $P<0.05$ compared to wildtype Rck).

The GGXWRDDXK motif is not sufficient to aggregate but mediates heterologous autoaggregation when surface exposed

It was of interest to determine whether the GGXWRDDXK motif mediated self-association or whether it was involved in some interactions with other parts of the Hra1 and Rck molecules. To address this question we firstly investigated whether a peptide containing the motif would aggregate. We synthesized the 10 amino acid sequence from Hra1, GGYWRDDLKE. Circular dichroism spectra of the resulting peptide was consistent with the peptide being largely unfolded and the wide peak of the IR spectrum of a 10 mm solution of the peptide suggests that it is most likely disorganized, existing in many conformations (Fig. S2). Sedimentation velocity data suggest that Hra1-derived peptide is mostly monomeric in solution (Fig. S3). Incubation of *E. coli* expressing Hra1 with the peptide at 100 mm, 10 mm and 1 mm did not produce any effect on the degree or rate of autoaggregation (data not shown). Taken together these findings suggest that the GGYWRDDLKE peptide is not sufficient to aggregate. To determine whether the GGXWRDDXK motif mediates self-association when in the context of an exposed loop on the cell surface, we performed a series of autoaggregation assays using mixed cultures. As shown in Fig. 6, Rck, like Hra1, is a self-associating autoaggregant. The degree of autoaggregation seen in 1 : 1 mixtures of *E. coli* cultures expressing Hra1 with cultures expressing Rck strongly suggests that Rck and Hra1 associate with one another. When the GGXWRDDXK motif is removed from either protein, it does not associate with its wildtype version or with the wildtype version of the other protein carrying the motif (Fig. 6). These data point to the GGXWRDDXK motif as a self-associating one and suggest that it promotes autoaggregation between heterologous bacteria, provided that each of the aggregating strains presents GGXWRDDXK on a surface-exposed loop.

DISCUSSION

Autoaggregation contributes to colonization directly by allowing entering bacteria to latch onto established bacteria in a niche, as well as indirectly by protecting bacterial cells from harmful proteins and phagocytic cells (Danese *et al.*, 2000). Autoaggregation has been reported to heighten virulence (Frick *et al.*, 2000) and improve transmission of bacterial pathogens. It may also be crucial for optimal dynamics in biofilms (Herman-Bausier *et al.*, 2015). Most mucosal colonizers can exhibit autoaggregation *in vitro* and/or carry genes that can confer this phenotype.

Autoaggregation by *E. coli* can be mediated by a variety of surface factors, including pili. However, non-structural self-associating surface proteins may actually produce more intimate associations because of their proximity to the cell surface (Frick *et al.*, 2000; Sherlock *et al.*, 2005). Included in this description are the multifunctional SAAT proteins, conferring, like the agglutinins, adherence to eukaryotic cells, surface adherence and/or invasion. SAATs produce *in vitro* autoaggregation at a much more rapid rate than Hra1. SAAT proteins are much larger than Hra1 and their functional regions (the autotransporter passenger domain 499 residues for Agn43, 798 for AIDA-1) probably dangle much further away from the cell than the comparatively short (25–35 residue) loops of agglutinins. Therefore, although Hra1 is slower to autoaggregate, it is probable that it produces tight aggregates.

Associations of large hydrophobic domains of bacterial surface proteins can produce autoaggregation (Menozzi *et al.*, 1994). Additionally, self-association has been known to be the result of coiled-coil dimers forming between molecules with sufficient extension on adjacent cells (Frick *et al.*, 2000; Nilson *et al.*, 1995; Phillips *et al.*, 1981). There are multiple *agn43* alleles and some of them are unable to demonstrate autoaggregation. By swapping portions of autoaggregating alleles with non-aggregating ones, Klemm *et al.* (2004) found that the autoaggregating phenotype of the Agn43 protein is conferred by residues in the first 160 amino acids of its externally-exposed 499-residue passenger domain. Heras *et al.* (2014) subsequently showed that a head-to-tail Velcro-like interaction between the inner long arm and the outer edge, of what they discovered to be L-shaped passenger domains of antigen 43 on adjacent cells, is the molecular explanation for autoagregation (Heras *et al.*, 2014). When Meng *et al.* (2011) solved the structure of the passenger domain of *Haemophilus influenzae* SAAT, they revealed it to be a prism produced from a β-strand coil with three β-strand faces. They were able to demonstrate that autoaggregation in this protein is produced by interactions between one β-strand 'face' of the triangle and one of the between-face edges. These few examples reveal that autoaggregation mechanisms can be simple or sophisticated and that different proteins can self-associate via a variety of means.

Structure–function studies have identified adhesin motifs that interact with eukaryotic cell surface receptors (Cirillo *et al.*, 1996; Fleckenstein *et al.*, 2002), but no autoaggregation motifs have been reported previously for self-associating integral outer membrane proteins. The surface-exposed functional domains of such proteins share little similarity (Sherlock *et al.*, 2004). Their high level of overall similarity arises from conserved β-strands that are lodged within the membrane and provide the structural scaffold for functional loops. The predicted surface-exposed loops of Hra1 are largely hydrophilic, lack coiled-coils and are comparatively short. Therefore interactive mechanisms different from those described in SAATs and coiled-coil proteins are likely to be at play. These mechanisms probably include, but are

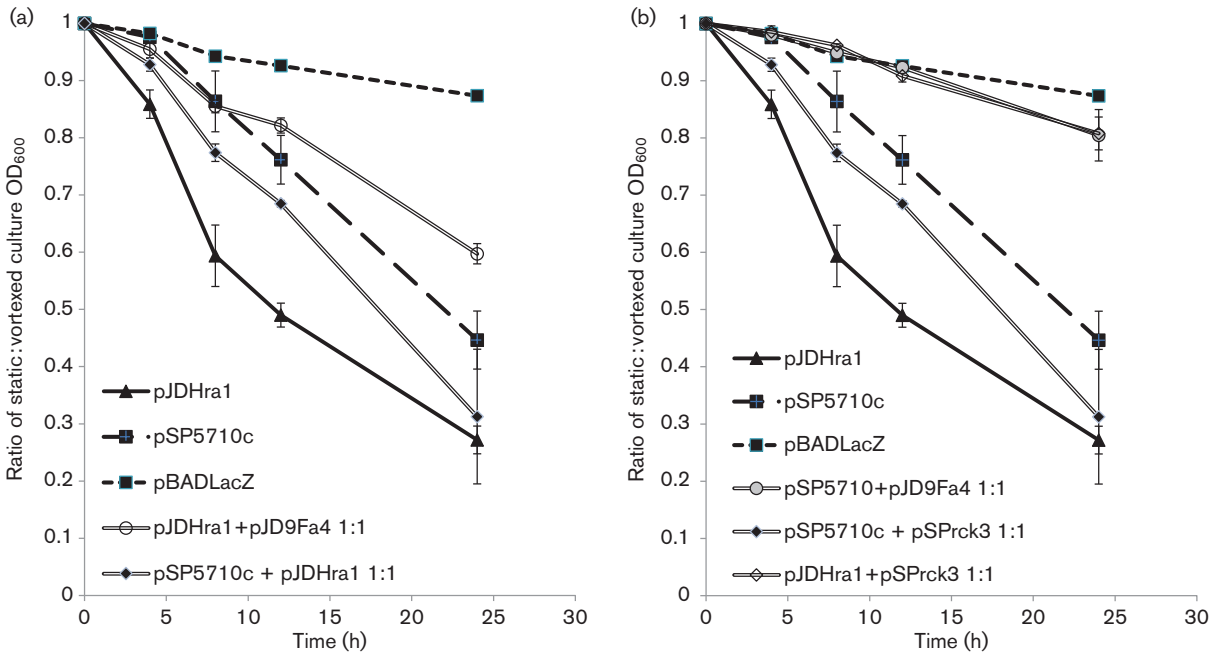

Fig. 6. Rck, like Hra1, mediates autoaggregation by self-association. Mixed cultures (1 : 1) of bacteria expressing Rck (pSP5710) and LacZ (pBADLacZ) from the arabinose promoter autoaggregate half as much as cultures containing bacteria expressing Rck alone. Mixed cultures (1 : 1) of bacteria expressing Rck (pSP5710) and Hra1 (pJDHra1) from the arabinose promoter autoaggregate to a degree intermediate between the autoaggregation seen with Hra1 alone and Rck alone, suggestive of a Rck–Hra1 interaction. (a) Removal of the GGXWRDDXK motif from Hra1 (pJG9F4) obliterates Hra1–Rck association. (b) Removal of the GGXWRDDXK motif from Rck (pSPrck3) obliterates Rck self-association and 1 : 1 mixed cultures of bacteria expressing Rck and Hra1 mutants that both lack the GGXWRDDXK motif in both Hra1 and Rck do not autoaggregate.

not restricted to, electrostatic interactions. In this study, we use Hra1, which confers autoaggregation by self-association, and its close homologue Tia, which does not, as the basis for finding autoaggregation motifs.

We recently described a third *E. coli* agglutinin, Hra2. As we reported previously, Hra2, like Tia, does not confer autoaggregation (Mancini *et al.*, 2011). Of the nine motifs/residues swapped in this study Hra2 is more similar to Tia in three, and to in Hra1 in three others (Fig. 3). The final three are unique to Hra2. Importantly, Hra2 is most similar to Tia in loop2 and most similar to Hra1 in loop 3 (including Hra1's Ser178) (Mancini *et al.*, 2011). It is, therefore, possible to predict from sequence comparisons that Hra2 would be unlikely to show significant autoaggregation and this is indeed the case (Mancini *et al.*, 2011).

The predicted surface loops of the agglutinin proteins are enriched for hydrophilic and charged residues and it is probable that charge-mediated interactions are responsible for at least some of the self-association conferred by some of the Hra1 motifs identified in this study. It was, therefore, of value to determine whether motifs that we found in the agglutinin family might confer autoaggregation on unrelated proteins as evidence that a more specific interaction might confer autoaggregation. Altogether we identified three such motifs and, while two of them were too short to reliably seek them in other contexts, the longest motif, GGXWRDDXK, is present in at least two proteins that could be surface exposed in bacteria. In the present study we have been able to determine that in the outer surface protein Rck, this motif confers autoaggregation by self-association.

Cirillo *et al.* (1996) used random mutagenesis to identify the Rck residues required for serum resistance and invasion. These authors found that Gyr118 and Asp43, both of which are conserved in Ail, are required for these phenotypes. These residues map to the third and first predicted external loops of Rck, respectively. Although *rck* has been extensively studied for serum resistance and invasion, a role in autoaggregation has not been reported. This is not surprising, given that Rck confers only weak autoaggregation in a heterologous background. Residues 80–89 of Rck are GGMSWRDDVK, the Hra1 motif that we located in the sequence and are predicted to lie within the second of four surface-exposed loops of that protein, a region not containing residues required for invasion or serum resistance. The motif is absent from Rck homologues OmpX, Ail, and Lom as well as from PagC alleles from *Salmonella* and *Pantoea*.

The GGXWRDDXK motif identified in the present study confers autoaggregation on Hra1 as well as on the unrelated

outer membrane protein, Rck. Interestingly, this motif also appears to mediate Hra1–Rck association. Other outer membrane proteins that produce autoaggregation in *E. coli* include the SAAT proteins: Antigen 43, TibA, AIDA-1 and Cah (Klemm *et al.*, 2004; Sherlock *et al.*, 2004, 2005; Torres *et al.*, 2002). *E. coli* SAAT proteins AIDA-1 and Antigen 43 are able to associate with each other (Sherlock *et al.*, 2004). In contrast, Hra1 did not autoaggregate with Tia, although Tia does not significantly self-associate. The Hra1–Rck association could theoretically bring together related but different bacterial species that share the same ecological niche. As Rck is activated by acyl homoserine lactone-mediated quorum sensing (Liu *et al.*, 2014) and Hra1 is commonly expressed by exceptional commensal colonizers as well as pathogens (Mancini *et al.*, 2011), the Rck–Hra1 interaction provides a possible, if untested, mechanism for niche invasion by *Salmonella* in the intestine. Provided autoaggregating and non-autoaggregating protein pairs can be identified within families, the methodology we have used here makes it possible to identify other motifs. In addition to the potential of being able to propose possible functions for unknown proteins that carry the motif, the existence of functional sequences, such as the GGXWRDDXK, offer the attractive proposition of blocking autoaggregation among very different organisms by interfering with a common motif.

ACKNOWLEDGEMENT

This work was supported by NSF RUI awards 0516591, 0948460 and 1329248 to INO. LIRF was supported by an undergraduate research award from the Howard Hughes Medical Institute through Haverford College. Strain SF1005 was a gift from Ferric C Fang and we thank the Korn Laboratory for the gift of a FLAG-tagged myosin clone. We are grateful to Justin Dorff and Erin C Remaly for technical assistance. We also thank Casey Londergan for the use of his FTIR instrument, and the Department Biochemistry and Biophysics at the University of Pennsylvania for the use of the MALDI-TOF mass spectrometer.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Barondess, J. J. & Beckwith, J. (1990). A bacterial virulence determinant encoded by lysogenic coliphage lambda. *Nature* **346**, 871–874.

Bhargava, S., Johnson, B. B., Hwang, J., Harris, T. A., George, A. S., Muir, A., Dorff, J. & Okeke, I. N. (2009). Heat-resistant agglutinin 1 is an accessory enteroaggregative *Escherichia coli* colonization factor. *J Bacteriol* **191**, 4934–4942.

Chart, H., Spencer, J., Smith, H. R. & Rowe, B. (1997). Identification of entero-aggregative *Escherichia coli* based on surface properties. *J Appl Microbiol* **83**, 712–717.

Chaudhuri, R. R., Sebaihia, M., Hobman, J. L., Webber, M. A., Leyton, D. L., Goldberg, M. D., Cunningham, A. F., Scott-Tucker, A., Ferguson, P. R. & other authors (2010). Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS One* **5**, e8801.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**, 3497–3500.

Cirillo, D. M., Heffernan, E. J., Wu, L., Harwood, J., Fierer, J. & Guiney, D. G. (1996). Identification of a domain in Rck, a product of the *Salmonella typhimurium* virulence plasmid, required for both serum resistance and cell invasion. *Infect Immun* **64**, 2019–2023.

Cooke, N. M., Smith, S. G., Kelleher, M. & Rogers, T. R. (2010). Major differences exist in frequencies of virulence factors and multidrug resistance between community and nosocomial *Escherichia coli* bloodstream isolates. *J Clin Microbiol* **48**, 1099–1104.

Czeczulin, J. R., Whittam, T. S., Henderson, I. R., Navarro-Garcia, F. & Nataro, J. P. (1999). Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. *Infect Immun* **67**, 2692–2699.

Dam, J., Velikovsky, C. A., Mariuzza, R. A., Urbanke, C. & Schuck, P. (2005). Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions c(s). *Biophys J* **89**, 619–634.

Danese, P. N., Pratt, L. A. & Kolter, R. (2000). Exopolysaccharide production is required for development of *Escherichia coli* K-12 biofilm architecture. *J Bacteriol* **182**, 3593–3596.

Deng, Q., Luo, W. & Donnenberg, M. S. (2007). Rapid site-directed domain scanning mutagenesis of enteropathogenic *Escherichia coli* espD. *Biol Proced Online* **9**, 18–26.

Evans, D. G., Silver, R. P., Evans, D. J., Chase, D. G. & Gorbach, S. L. (1975). Plasmid-controlled colonization factor associated with virulence in Esherichia coli enterotoxigenic for humans. *Infect Immun* **12**, 656–667.

Fagan, R. P. & Smith, S. G. (2007). The Hek outer membrane protein of *Escherichia coli* is an auto-aggregating adhesin and invasin. *FEMS Microbiol Lett* **269**, 248–255.

Fagan, R. P., Lambert, M. A. & Smith, S. G. (2008). The Hek outer membrane protein of *Escherichia coli* strain RS218 binds to proteoglycan and utilizes a single extracellular loop for adherence, invasion, and autoaggregation. *Infect Immun* **76**, 1135–1142.

Fang, F. C., Libby, S. J., Buchmeier, N. A., Loewen, P. C., Switala, J., Harwood, J. & Guiney, D. G. (1992). The alternative sigma factor katF (rpoS) regulates Salmonella virulence. *Proc Natl Acad Sci U S A* **89**, 11978–11982.

Fleckenstein, J. M., Kopecko, D. J., Warren, R. L. & Elsinghorst, E. A. (1996). Molecular characterization of the tia invasion locus from enterotoxigenic *Escherichia coli*. *Infect Immun* **64**, 2256–2265.

Fleckenstein, J. M., Holland, J. T. & Hasty, D. L. (2002). Interaction of an uuter membrane protein of enterotoxigenic *Escherichia coli* with cell surface heparan sulfate proteoglycans. *Infect Immun* **70**, 1530–1537.

Frick, I. M., Mörgelin, M. & Björck, L. (2000). Virulent aggregates of *Streptococcus pyogenes* are generated by homophilic protein-protein interactions. *Mol Microbiol* **37**, 1232–1247.

Hasman, H., Chakraborty, T. & Klemm, P. (1999). Antigen-43-mediated autoaggregation of *Escherichia coli* is blocked by fimbriation. *J Bacteriol* **181**, 4834–4841.

Heffernan, E. J., Harwood, J., Fierer, J. & Guiney, D. (1992a). The *Salmonella typhimurium* virulence plasmid complement resistance gene rck is homologous to a family of virulence-related outer membrane protein genes, including pagC and ail. *J Bacteriol* **174**, 84–91.

Heffernan, E. J., Reed, S., Hackett, J., Fierer, J., Roudier, C. & Guiney, D. (1992b). Mechanism of resistance to complement-mediated killing of bacteria encoded by the *Salmonella typhimurium* virulence plasmid gene rck. *J Clin Invest* **90**, 953–964.

Heras, B., Totsika, M., Peters, K. M., Paxman, J. J., Gee, C. L., Jarrott, R. J., Perugini, M. A., Whitten, A. E. & Schembri, M. A. (2014).

The antigen 43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping. *Proc Natl Acad Sci U S A* **111**, 457–462.

Herman-Bausier, P., El-Kirat-Chatel, S., Foster, T. J., Geoghegan, J. A. & Dufrêne, Y. F. (2015). *Staphylococcus aureus* fibronectin-binding protein A mediates cell-cell adhesion through low-affinity homophilic bonds. *MBio* **6**, e00413–00415.

Klemm, P., Hjerrild, L., Gjermansen, M. & Schembri, M. A. (2004). Structure-function analysis of the self-recognizing Antigen 43 autotransporter protein from *Escherichia coli*. *Mol Microbiol* **51**, 283–296.

Klemm, P., Vejborg, R. M. & Sherlock, O. (2006). Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol* **296**, 187–195.

Kolodziejek, A. M., Schnider, D. R., Rohde, H. N., Wojtowicz, A. J., Bohach, G. A., Minnich, S. A. & Hovde, C. J. (2010). Outer membrane protein X (Ail) contributes to *Yersinia pestis* virulence in pneumonic plague and its activity is dependent on the lipopolysaccharide core length. *Infect Immun* **78**, 5233–5243.

Liu, Z., Que, F., Liao, L., Zhou, M., You, L., Zhao, Q., Li, Y., Niu, H., Wu, S. & other authors (2014). Study on the promotion of bacterial biofilm formation by a *Salmonella* conjugative plasmid and the underlying mechanism. *PLoS One* **9**, e109808.

Lutwyche, P., Rupps, R., Cavanagh, J., Warren, R. A. & Brooks, D. E. (1994). Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglutinin 1, an integral membrane hemagglutinin from *Escherichia coli* O9:H10:K99. *Infect Immun* **62**, 5020–5026.

Lutwyche, P., Norris-Jones, R. & Brooks, D. E. (1995). Aqueous two-phase polymer systems as tools for the study of a recombinant surface-expressed *Escherichia coli* hemagglutinin. *Appl Environ Microbiol* **61**, 3251–3255.

Mammarappallil, J. G. & Elsinghorst, E. A. (2000). Epithelial cell adherence mediated by the enterotoxigenic *Escherichia coli* Tia protein. *Infect Immun* **68**, 6595–6601.

Mancini, J., Weckselblatt, B., Chung, Y. K., Durante, J. C., Andelman, S., Glaubman, J., Dorff, J. D., Bhargava, S., Lijek, R. S. & other authors (2011). The heat-resistant agglutinin family includes a novel adhesin from enteroaggregative *Escherichia coli* strain 60A. *J Bacteriol* **193**, 4813–4820.

Meng, G., Spahich, N., Kenjale, R., Waksman, G. & St Geme, J. W., 3rd. (2011). Crystal structure of the *Haemophilus influenzae* Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. *EMBO J* **30**, 3864–3874.

Menozzi, F. D., Boucher, P. E., Riveau, G., Gantiez, C. & Locht, C. (1994). Surface-associated filamentous hemagglutinin induces autoagglutination of *Bordetella pertussis*. *Infect Immun* **62**, 4261–4269.

Nataro, J. P., Scaletsky, I. C., Kaper, J. B., Levine, M. M. & Trabulsi, L. R. (1985). Plasmid-mediated factors conferring diffuse and localized adherence of enteropathogenic *Escherichia coli*. *Infect Immun* **48**, 378–383.

Nilson, B. H., Frick, I. M., Akesson, P., Forsén, S., Björck, L., Akerström, B. & Wikström, M. (1995). Structure and stability of protein H and the M1 protein from *Streptococcus pyogenes*. Implications for other surface proteins of Gram-positive bacteria. *Biochemistry* **34**, 13688–13698.

Phillips, G. N., Flicker, P. F., Cohen, C., Manjula, B. N. & Fischetti, V. A. (1981). Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc Natl Acad Sci U S A* **78**, 4689–4693.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sherlock, O., Schembri, M. A., Reisner, A. & Klemm, P. (2004). Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm formation. *J Bacteriol* **186**, 8058–8065.

Sherlock, O., Vejborg, R. M. & Klemm, P. (2005). The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. *Infect Immun* **73**, 1954–1963.

Söding, J., Biegert, A. & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244–248.

Srinivasan, U., Foxman, B. & Marrs, C. F. (2003). Identification of a gene encoding heat-resistant agglutinin in *Escherichia coli* as a putative virulence factor in urinary tract infection. *J Clin Microbiol* **41**, 285–289.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.

Torres, A. G., Perna, N. T., Burland, V., Ruknudin, A., Blattner, F. R. & Kaper, J. B. (2002). Characterization of Cah, a calcium-binding and heat-extractable autotransporter protein of enterohaemorrhagic *Escherichia coli*. *Mol Microbiol* **45**, 951–966.

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M. & other authors (2014). PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res* **42**, W337–W343.

Zolkiewski, M., Redowicz, M. J., Korn, E. D., Hammer, J. A. & Ginsburg, A. (1997). Two-state thermal unfolding of a long dimeric coiled-coil: the *Acanthamoeba myosin* II rod. *Biochemistry* **36**, 7876–7883.

Edited by: M. Whiteley