molecular microbiology

The passenger-associated translocation repeat promotes virulence factor secretion efficiency and delineates a distinct autotransporter subtype.

Journal:	*Molecular Microbiology*
Manuscript ID:	MMI-2015-14886
Manuscript Type:	Research Article
Date Submitted by the Author:	11-Jan-2015
Complete List of Authors:	Doyle, Matthew; The University of Adelaide, Molecular and Cellular Biology Tran, Elizabeth; The University of Adelaide, Molecular and Cellular Biology Morona, Renato; The University of Adelaide, Molecular and Cellular Biology
Key Words:	Autotransporter, Secretion, Virulence Factor, Repeat, Outer Membrane

TITLE:

The passenger-associated translocation repeat promotes virulence factor secretion efficiency and delineates a distinct autotransporter subtype.

RUNNING TITLE:

PATR directed secretion of autotransporters.

AUTHORS:

Matthew Thomas Doyle, Elizabeth Ngoc Hoa Tran, and Renato Morona

ADDRESS WORK WAS PERFORMED:

Department of Molecular and Cellular Biology, School of Biological Sciences, University of Adelaide, Adelaide 5005, Australia

(Formerly Discipline of Microbiology and Immunology, School of Molecular and Biomedical Science)

CORRESPONDING EMAIL: Renato Morona: renato.morona@adelaide.edu.au.

Tel: +61 8 8313 4151 Fax: +61 8 8313 7532

KEY WORDS:

Autotransporter, Secretion, Virulence Factor, Repeat, Outer Membrane, Translocation

SUMMARY

Autotransporters are a superfamily of virulence factors secreted by Gram negative bacteria. They are comprised of an N-terminal passenger domain that is translocated across the outer membrane, and a C-terminal domain that inserts into the outer membrane forming a β-barrel anchor. It is still poorly understood how the passenger is efficiently translocated in the absence of external energy inputs. Several mechanisms have been proposed in solution of this problem, yet due to the vast diversity of size, sequence, and function of the passenger, it is not clear how widely these mechanisms are employed. In this study we functionally characterize a conserved repeat found in many passengers which we designate the Passenger-associated Translocation Repeat (PATR). Using the autotransporter IcsA from the enteropathogen *Shigella flexneri*, we identified conserved PATR residues that are required for efficient translocation of the passenger during growth and infection. Furthermore, PATR-containing autotransporters are significantly larger than non-PATR autotransporters, with PATR copy number correlating with passenger size. We also show that PATR-containing autotransporters delineate a subgroup that associates with specific virulence traits. These results advance our understanding of autotransporter composition, and indicate that an additional modular mechanism of passenger translocation is in use in thousands of these proteins.

INTRODUCTION

Gram negative bacteria coordinate infection and disease via a synergy of secreted virulence factors. As such, the secretion of these virulence factors across the double-membrane cell wall must be efficient for pathogenic fitness. The autotransporter (AT, or Type Va) secretion pathway is the most common solution to this problem (Henderson *et al.*, 2004; Dautin *et al.*, 2007; Leyton *et al.*, 2012; Grijpstra *et al.*, 2013). AT superfamily proteins have distinctive domain architecture; (i) an N-terminal signal sequence for Sec mediated passage of the inner membrane, (ii) a central passenger domain harboring the virulence properties of the protein, and (iii) a transmembrane β-barrel at the C-terminus. The β-barrel is required for the stages of outer membrane (OM) insertion, and the translocation of the passenger to the extracellular space. Depending on the virulence function of the AT, the passenger can remain attached to the bacterial surface or be released into the extra-bacterial milieu. Also, due to its functional diversity the passenger varies widely in its sequence and size (Celik *et al.*, 2012).

Despite rigorous investigation, the exact mechanism of OM β-barrel insertion and passenger translocation remains incompletely understood. In general, the β-barrel and passenger are inserted and translocated sequentially by a series of intricate events coordinated by the essential Barrel Assembly Machinery (BAM) (Jain *et al.*, 2007; Rossiter *et al.*, 2011; Roman-Hernandez *et al.*, 2014) and by the β-barrel itself (Pavlova *et al.*, 2013; Leyton *et al.*, 2014). The BAM (a complex of integral and lipoproteins) interacts with both the β-barrel (Ieva *et al.*, 2011) and passenger (Ieva *et al.*, 2009) portions of the AT to assist β-barrel insertion. Structural analysis suggests that the local distortion of the OM at the lateral gate of the BamA component facilitates the seeding of the nascent β-barrel into the membrane (Noinaj *et al.*, 2013; Noinaj *et al.*, 2014).

The passenger is then translocated in a C- to N-terminal manner though the nascent β-barrel pore where it commonly folds into a β-helical stalk (Kajava *et al.*, 2006).

The energy source for passenger translocation also remains under debate. In the absence of external energy sources such as ATP, it has been proposed that C- to N-terminal β-helical folding could drive passenger translocation (Junker *et al.*, 2006; Braselmann *et al.*, 2012). In this model, the initially translocated C-terminal portion of the passenger has high stability and folds with high efficiency (Junker *et al.*, 2006; Renn *et al.*, 2008; Peterson *et al.*, 2010). This folding action itself pulls the rest of the less stable N-terminal portion of the passenger through the β-barrel pore in a vectorial fashion. This type of sequential folding, although attractive in its simplicity, has only been implicated in a few model ATs (such as Petactin, EspP, and Pet (Renn *et al.*, 2008; Junker *et al.*, 2009; Peterson *et al.*, 2010; Kang'ethe *et al.*, 2013b)) and it is unknown how widely this model is adhered to by the AT superfamily which is large and diverse. Indeed, the AT YapV varies from the model as it lacks a C-terminal (Pertactin-like, PL) stable region but employs an unstable region at the extreme N-terminus for its secretion (Besingi *et al.*, 2013). Furthermore, it has been recognized that the vast majority of ATs do not contain a PL region (Drobnak *et al.*, 2014) and the model may not apply to ATs that possess globular passengers (for instance EstA (van den Berg, 2010)). It has also been realized that passenger domains frequently have a net negative charge that may act to facilitate translocation, possibly by charge repulsion between the passenger and LPS molecules (Kang'ethe *et al.*, 2013a). Consequently, it remains possible that mechanisms of passenger secretion are 'mixed-and-matched' depending on the size, fold-type, or function of the AT in question.

To further our understanding of AT passenger secretion, we investigated a highly conserved, AT-associated, 32 amino acid repeat. Although models of this repeat had been deposited into

both TIGRFAM (Haft *et al.*, 2003) (TIGR02601) and Pfam (Finn *et al.*, 2014) (PF12951) databases, its function remained completely uncharacterized. This study revealed that the repeat is; (i) associated with the passengers of a large and district group of ATs, (ii) required for efficient passenger translocation, and (iii) connected with certain passenger domain architectures and functions. As such, in this study we refer to this repeat as the Passenger-associated Translocation Repeat (PATR).

RESULTS

PATR mutation disrupts steady-state passenger surface presentation.

To uncover the function of the PATR in ATs, the well-studied AT IcsA was used as a model protein. IcsA is an essential virulence factor for the human enteric pathogen *Shigella flexneri* that enables spreading and lesion formation in infected intestinal epithelia (Makino *et al.*, 1986; Bernardini *et al.*, 1989; Lett *et al.*, 1989; Goldberg *et al.*, 1995; Suzuki *et al.*, 2002). It also has a previously reported (Dai *et al.*, 2010), but uninvestigated, single copy of the PATR in its passenger domain at IcsA$^{526\text{-}557}$ (see schematic Figure 1A). On inspection of the Pfam (PF12951) PATR model (see Figure 1B), the occurrence of four highly conserved glycines was observed at positions G^{6}, G^{8}, G^{20}, and G^{27} that are present in IcsA (IcsA^{G531}, IcsA^{G533}, IcsA^{G545}, and IcsA^{G552} respectively). We hypothesized that the PATR glycine residues are important in biogenesis, and accordingly, glycine to alanine substitutions at these four sites were constructed in IcsA, along with a complete 32 amino acid deletion of the PATR. These mutants were constructed in a plasmid-borne *icsA* with a native P_{IcsA} promoter (Morona *et al.*, 2003c) and expressed in an IcsA and O-antigen deficient strain of *S. flexneri* (Van den Bosch *et al.*, 2003) (see Table S1). This

allowed unhindered detection of IcsA surface levels (Morona *et al.*, 2003a; Morona *et al.*, 2003c; Morona *et al.*, 2003b) making strains suitable for the following experiments (see further).

First, to assess overall IcsA PATR mutant expression levels, total *S. flexneri* protein samples were analyzed by anti-IcsA (passenger) Western immunoblot (Figure 2A). No difference in total IcsA protein expression was observed, regardless of PATR mutation. Anti-IcsA immunofluorescence (IF) staining of these bacteria was also conducted (Figure 2B). Contrary to total protein levels, detectable passenger surface levels were visually reduced for IcsA PATR mutants relative to the wild-type protein. This was particularly true for IcsA^{G531A}, IcsA^{G545A}, IcsA^{G552A} substitutions, and IcsA$^{\Delta PATR}$. IcsA fluorescence intensities of 250 bacteria (n = 5) were then measured for each PATR mutant and the wild-type protein (Figure 2C). This confirmed that IcsA^{G531A}, IcsA^{G545A}, IcsA^{G552A}, and IcsA$^{\Delta PATR}$ displayed significantly ($\alpha = 0.05$) reduced surface levels (mean fractions of wild-type 0.58 ±0.20, 0.61 ±0.31, 0.61 ±0.17, and 0.62 ±0.08 respectively). Although this experiment could not establish significance for IcsA^{G533A} ($P = 0.588$), the trend of reduced surface detection was still observable.

The reduction of detectable surface exposed passenger due to PATR glycine substitutions or deletion may have been due to a reduction of biogenesis towards the OM. To test this, OM protein (OMP) samples were extracted from *S. flexneri* expressing IcsA and IcsA PATR mutants using differential detergent treatment (see *Materials and Methods*). Coomassie Blue staining of SDS-PAGE separated OMP samples showed equivalent loading and excellent enrichment of major OMPs relative to a total cell protein control (Figure 2D). OMP enrichment was further shown by anti-BamA immunoblotting. Fractionation purity controls showed that periplasmic SurA, inner membrane Wzz, and cytoplasmic DnaK did not contaminate the OMP samples. Interestingly, there was no discernable difference in the amount of full length IcsA in the OM for

any of the IcsA PATR mutants relative to the wild-type protein. What was noticeable however, was an increase in OM-associated IcsA degradation products due to PATR mutation. This was observed for all IcsA PATR substitutions and deletion, but most clearly observed for IcsA^{G545A}.

The reduction in detectable IcsA on the *S. flexneri* surface (Figures 2B and C) despite equivalent cellular expression and OM localization (Figures 2A and D) indicated a defect in efficient passenger translocation due to either PATR glycine substitutions or deletion. This would also explain the increase in the degraded forms of IcsA in the OM (Figure 2D). If IcsA passenger translocation was reduced due to PATR glycine substitutions or deletion, then this would also be observed by a reduction of N-WASP recruitment by the bacterium upon infection of epithelial cells. Recruitment of host N-WASP is the intracellular function of IcsA that results in the formation of actin-based tails required for motility and pathogenicity (Egile *et al.*, 1999; Suzuki *et al.*, 2002). To test this, cultured HeLa cells were infected with *S. flexneri* expressing either the wild-type or PATR mutant forms of IcsA and stained for N-WASP, filamentous-actin, and DNA. As expected, wild-type IcsA recruited high levels of N-WASP resulting in commonly observed tail filaments (Figure 2E). However, dramatic reductions in N-WASP recruitment was observed for substitutions IcsA^{G531A}, IcsA^{G533A}, IcsA^{G545A}, and IcsA^{G552A}, and was completely abolished for IcsA$^{\Delta PATR}$ indicating that portions of the passenger N-terminal to the deletion were completely misfolded. To confirm these observations, the N-WASP fluorescence intensity was measured for all bacteria per infected cell (n = 5) for each strain expressing the PATR mutants and the wild-type protein (Figure 2F). This confirmed that IcsA^{G531A}, IcsA^{G533A}, IcsA^{G545A}, IcsA^{G552A}, and IcsA$^{\Delta PATR}$ all resulted in significant ($\alpha = 0.05$) reductions in N-WASP recruitment (mean fractions of wild-type 0.50 ±0.12, 0.51 ±0.10, 0.35 ±0.06, 0.50 ±0.07, and 0.22 ±0.02

158 respectively). Also, N-WASP recruitment for IcsA^{G545A} and IcsA$^{\Delta PATR}$ was not significantly
159 different to the vector control ($p = 0.174$ and $p = 0.873$ respectively).
160
161 **PATR mutation decreases passenger translocation efficiency.**
162 Thus far, a role for the PATR in steady-state passenger surface presentation, in the context of
163 both bacterial culture and infection, had been established. To more closely investigate the
164 dynamics of IcsA passenger translocation, the same IcsA PATR glycine substitutions and
165 deletion were constructed in the plasmid pBADIcsA which has the *icsA* gene controlled by the
166 P_{BAD} promoter (Guzman *et al.*, 1995). Expression of these constructs in an IcsP protease
167 deficient strain of *S. flexneri* (see Table S1) allowed reduced endogenous proteolysis permitting
168 informative pulse-chase protease accessibility assays to be performed (Figure 3). Briefly, IcsA
169 expression was momentarily pulsed by the addition of arabinose in culture, and the newly
170 synthesized protein was chased by sampling over an hour time-course (see *Materials and*
171 *Methods*). *S. flexneri* samples were either untreated (PK-) or treated with Proteinase K (PK+),
172 allowing assessment of passenger translocation rate due to protease accessibility. As a further
173 control, samples were treated with chloroform to permeabilize the OM and allow periplasmic
174 access to PK. Correct topological proteolysis was established in a mock pulse-chase (Figure 3A).
175 This confirmed that digestion of periplasmic SurA only occurred after OM permeabilization.
176 Additionally, the cytoplasmic protein DnaK was not affected by any treatment indicating that the
177 inner membrane remained impermeable to PK under the conditions of this experiment.
178 The protease accessibility chase for wild-type IcsA revealed an extremely fast rate of
179 passenger translocation (Figure 3B and C, black trace) with a 'burst' of 93.10% (±2.0)
180 translocation occurring in the first 1 to 5 minutes with the remainder still translocating (as

indicated by complete digestion after OM permeabilization). Wild-type IcsA translocation was totally completed between 20 to 40 minutes. This translocation rate is expected since it is consistent with the life-cycle of rapidly dividing *S. flexneri* that require sufficient surface exposed IcsA passenger in order to initiate motility and pathogenic viability in the host. The IcsA PATR glycine substitutions and deletion mutants however, displayed marked reductions in passenger translocation efficiency (Figure 3B and C). For instance, the initial bursts of translocation between the first 1 to 5 minutes were reduced by approximately half for IcsA^{G531A}, IcsA^{G533A}, and IcsA^{G552A}, (27.80% ±6.7, 41.12% ±35.3, and 31.37% ±12.6 respectively) and 59.39% ±25.73 for IcsA^{G545A}. Furthermore all PATR glycine substitutions had increased their translocation percentage over the hour time-course, but none progressed to complete translocation. Interestingly, the PATR deletion mutant had decreased translocation by the first 5 minutes (63.51% ±24.4), but also declined in translocation across the time-course. This may indicate a complete blockage in translocation where additional nascent IcsA$^{\Delta PATR}$ is immediately blocked at the translocation stage with an accumulative effect. Finally, the mean translocation percentages over all time points were analyzed by repeated measures ANOVA and were found to be significantly lower than the wild-type passenger (mean differences of 66.59% ±7.3, 41.38% ±9.2, 35.16% ±9.5, 52.23% ±8.3, and 65.11% ±14.1 relative wild-type for IcsA^{G531A}, IcsA^{G533A}, IcsA^{G545A}, IcsA^{G552A}, and IcsA$^{\Delta PATR}$ respectively) (Figure 3D).

These results show that PATR glycine substitutions or deletion caused a marked decrease (approximately halved) in passenger translocation over the initial 5 minutes and a continual lag in translocation thereafter. This provides an explanation for the reduced levels of surface exposed PATR mutant passengers observed for the steady-state circumstances (Figure 2).

Analysis of the PATR within the AT family.

With knowledge of the function of the PATR established *in vivo*, we attempted to identify the wider importance of the PATR within the AT family. The PATR was identified in a large number of ATs, with examples shown aligned in Figure 4A. These include known subtilisin-type serine proteases such as the inflammatory EprS from *Pseudomonas aeruginosa* (Kida *et al.*, 2013), and NalP, the processor of other ATs from *Neisseria meningitidis* (van Ulsen *et al.*, 2003). Also shown aligned is PATR9 from the fibronectin-binding host colonization factor ShdA of *Salmonella enterica* (ShdA contains an array of PATRs (Kingsley *et al.*, 2000; Kingsley *et al.*, 2004) (see Figure S1)). The glycine residues investigated in this study, as well as other PATR residues, are highly conserved on a level not previously observed in AT passengers, especially between those of varied function.

An analysis on all ATs within the UniProt Knowledgebase (7659 proteins) was also conducted by grouping unique ATs (InterPro (Hunter *et al.*, 2012) AT β-barrel identifiers IPR005546 and/or IPR006315) based on the presence or absence of at least one annotated copy of the PATR (IPR013425). Remarkably, 29.2% of the unique representative ATs within the database (2240 proteins) had at least one copy of the PATR. This is similar to the abundance of the PL region which we found present in 37.4% of the unique representative ATs (2864 proteins). There was a significant difference ($p < 0.0001$) in the representations of passenger domain virulence traits between PATR-positive and non-PATR ATs (Figure 4B). For example; lipase-like, as well as type S6 (SPATE-like (Rawlings *et al.*, 2014; Ruiz-Perez *et al.*, 2014)) serine proteases, and vacuolating cytotoxins were only present in non-PATR ATs. Conversely, ATs containing type 2 phosphatidic acid phosphatase (PAP2) domains, Polymorphic OM protein repeats (POMPs), and type S8 (subtilisin-type (Rawlings *et al.*, 2014)) serine proteases were all

highly represented in PATR-positive ATs. Interestingly, ATs with a Pectin lyase-like region (an indicator of further β-helical wrapping) as the only other identifying passenger feature, were represented more than twice as high in PATR-type ATs relative non-PATR ATs. The inverse of this was true for ATs that also contained a PL region. Furthermore, the PATR was never observed in the passenger with the PL as an exclusive partnership (i.e.: PATR plus PL only passenger). This minimal overlap between the PATR and the PL is further shown via Venn diagram (see figure S2). Also noticed was a significant difference ($P < 0.0001$) in protein lengths, where the mean length of PATR-positive ATs was 503 a.a (± 13) longer than non-PATR ATs. This is seen as a positively skewed lengths distribution for ATs containing a PATR (Figure 4C). Moreover, within the PATR-positive ATs there was a significant ($P < 0.0001$) correlation between increasing AT length and PATR copy number (Figure 4D).

Together, these data suggests that the presence or absence of a PATR strongly impacts the probability of containing certain passenger virulence functions, the potential size of the protein, and delineates an important sub-group of PATR-type ATs. Surprisingly, none of the ATs with solved passenger structures appeared to contain the PATR when scrutinized by sequence analyses. Therefore, the tertiary structure of the PATR consensus sequence was modeled using I-TASSER (Roy *et al.*, 2010; Xu *et al.*, 2011) and then structurally aligned to all the solved AT passengers using TM-Align (Zhang *et al.*, 2005). The PATR was predicted to form a right-handed β-helical triangular wedge with all PATR glycines clustered at each vertex (Figure 4E). Remarkably, upon alignment to passenger structures, we identified putatively degenerate PATR sites in Ag43, Hap, and IgA1P (see Figures 4F, S3, and Table S2 for additional sites). The alignment shows the positions of the conserved glycines characterized in this study. The 'velcro-like' Ag43 (Heras *et al.*, 2014) of pathogenic *E. coli* had the highest alignment score with

clustering of the glycines at the $G^{6}/G^{8}/G^{27}$ PATR apex (Ag43 glycines G^{499}, G^{501}, G^{520} respectively). Identification of degenerate PATRs indicate an even wider distribution of this site in ATs than can be identified by sequence motif recognition alone.

DISCUSSION

This study reveals the importance of a previously underappreciated and uncharacterized passenger feature which we have termed here, the Passenger-associated Translocation Repeat (PATR). Strikingly, ~30% of the unique AT representatives analyzed contain a PATR. This is comparable to the abundance of the Pertactin-like (PL) region which we found here to be ~37% (slightly more than previous estimates (Drobnak *et al.*, 2014)). Through alanine substitution of conserved PATR glycines (as well as a complete PATR deletion) within the PATR-type AT IcsA, we have shown *in vivo* that the PATR is required for efficient translocation of the passenger during the initial few minutes of secretion. This resulted in significantly decreased (approximately half that of wild-type) steady-state levels of surface exposed IcsA passenger which was further observed as a significant decrease in N-WASP recruitment levels by intracellular *S. flexneri*. Furthermore, PATR mutants also displayed substantial lags in translocation over an extended time, also suggesting that the passenger was exposed to the periplasmic topology for a prolonged period. This would increase the likelihood of proteolysis by known proteases (for instance, DegP (Jong *et al.*, 2007; Purdy *et al.*, 2007)) and may explain the observed increase in degraded forms of PATR mutants in the OM.

The knowledge that the PATR is required for efficient translocation of the passenger begs the question – by what mechanism? Structural modeling suggests that the PATR prescribes a minimal right-handed triangular β-helix with the conserved glycines clustered spatially at the

three vertices. We found a translocation deficiency effect arising from glycine-alanine substitutions at two of these three putative vertices. It is likely that these glycines are required for the flexibility of stable PATR folding at the corners, and that substitution disrupts the space requirements for this folding. Sequential folding as a mode of translocation is a well-studied notion in other ATs (Junker *et al.*, 2006; Renn *et al.*, 2008; Junker *et al.*, 2009; Peterson *et al.*, 2010). Drawing from this, it is conceivable that proper sequential folding of the PATR is also required for the translocation of PATR-type ATs, similarly to what is observed for Pertactin and PL ATs. Certainly, it is intriguing that the presence of a PL region corresponds with a lower potential for the presence of a PATR and *vice versa*. Further, the PATR was never observed with the PL as an exclusive partnership in the passenger – an additional region (such as the Pectin lyase-like region) must also be present for the PATR and PL to be in the same AT. Although care must be taken when attributing a folding function to PL regions (Drobnak *et al.*, 2014), it appears that both these regions have overlapping (but not equivalent) conserved functions in passenger biogenesis. This rationale needs to be investigated further through *in vitro* biophysical and structural characterization of the PATR before it is validated. However, we have observed that PATR-type ATs are significantly larger (by 503±13 a.a.) than non-PATR ATs, and that the size of the AT correlates with PATR copy number. We propose that the PATR acts as a dispersed module for folding stability where some larger ATs (for instance ShdA) may require multiple PATR modules for the efficient translocation of larger passengers.

Besides the PL region, there was a striking association between passenger virulence attributes and the presence or absence of the PATR. The PATR was not found in lipase-like passengers, which is consistent with our modeling of the PATR as a minimal β-helical fold (since lipases are not β-helix based (van den Berg, 2010)). Subtilistin-type (S8) serine proteases

were also commonly PATR-type ATs, whereas SPATE (serine protease ATs of the enterobacteriacea)-like (S6) serine proteases generally excluded the presence of a PATR. The reasons behind these observations are unknown, but we speculate that the export requirements of a passenger with S8-type regions may be more amenable to a PATR-type mechanism as opposed to the requirements for S6-type passengers. It is also possible that the difference in representation reflects variances in the usage of the PATR between the enterobacteriacea and other families. In support of this we also observed an increased association of the PATR with passengers containing the POMP repeat which is highly conserved in the Pmp adhesin ATs of the *Chlamydiaceae* (Henderson *et al.*, 2001). Together, these results implicate the PATR as a convenient building block in passengers, providing scaffolding for other functional regions. Indeed, it has been previously proposed that small sequences have been incorporated into passenger architectures to enable niche specialization of ATs during evolution (Celik *et al.*, 2012).

In conclusion, this study has uncovered the importance of a previously uncharacterized repeat that plays a role in AT secretion. The PATR delineates a further subtype of ATs and is present in many passengers. These results stimulate the need for further investigation to expose the exact mechanism of PATR mediated translocation, to establish biophysical characteristics of this repeat, and to uncover its phylogenic origin and diversity within Gram-negative bacteria. Finally, it should be noted here that we identified that the PATR was also present in ~700 unique proteins that did not contain an identifiable AT β-barrel. Although it must be further established, this may indicate a role for the PATR in other Type V secretion systems – namely, the Two-partner pathway. Nevertheless, this work has highlighted the notion that passenger compositions are ‘mixed-and-matched’ to suit precise secretion and function requirements.

EXPERIMENTAL PROCEDURES

Bacterial strains and plasmids. Lists of strains and plasmids utilized in this study are included in the Supporting Information (see Table S1) which includes details of their construction (see Text S1) and oligonucleotides used (see Table S3). *S. flexneri* colonies were grown on Congo Red agar for confirmation of virulence plasmid presence before routine growth in Luria-Bertani (LB) media at 37˚C with shaking. For all experiments, bacteria were sub-cultured (1:50 or 1:100) to a log-phase OD600 reading of 0.5 before use. When required, broths were supplemented with the following additives at respective concentrations; 0.2 % (w/v) glucose, 0.2 % (w/v) arabinose, tetracycline (10 µg/mL), kanamycin (50 µg mL^{-1}), chloramphenicol (25 µg mL^{-1}) and ampicillin (50 µg mL^{-1}).

Antibodies. Polyclonal rabbit anti-IcsA (passenger) and polyclonal rabbit anti-N-WASP were produced and validated as described previously (Van den Bosch *et al.*, 1997; May *et al.*, 2008). Polyclonal rabbit anti-SurA was a generous gift from Carol Gross (University of California, USA). Polyclonal rabbit anti-Wzz was produced as described previously (Daniels *et al.*, 1999). Polyclonal rabbit anti-BamA was a generous gift from Thomas Silhavy (Princeton University, USA). Mouse anti-DnaK monoclonal antibody was from Enzo Life Sciences.

Total bacterial protein samples. 5 x 10^8 of log-phase bacteria were collected by centrifugation (16000 x g, 1 min, 4 ˚C), resuspended in 100 µL of SDS-PAGE loading buffer (Lugtenberg *et al.*, 1975), and heated to 100 ˚C for 10 min. Replicate total cell samples were pooled 1:1 before analysis.

342

343 **Bacterial IcsA labeling.** Immunofluorescence microscopy and fluorescence quantitation was
344 conducted as described previously (Tran *et al.*, 2013). All solutions used were filtered through a
345 0.2 µm nitrocellulose filter. 10^8 of log-phase bacteria were harvested from a 1:50 sub-culture by
346 centrifugation (16000 x g, 2 min, 20 ˚C), resuspended in 3.7 % (v/v) formaldehyde solution
347 (Sigma) in phosphate buffered saline (PBS), and incubated at 20 ˚C for 20 min. Fixed bacteria
348 were washed twice in PBS before resuspension in 100 µL of PBS. 5 µL of the bacteria were
349 spotted onto sterile round coverslips (at the bottom of a 24-well tray) that were previously treated
350 with 10 % (v/v) poly-L-lysine solution (Sigma) in PBS. Bacteria were centrifuged (775 x g, 5
351 min, 20 ˚C) and then incubated for 2 h with anti-IcsA diluted 1:100 in PBS containing 10 % (v/v)
352 fetal calf serum (FCS). Bacteria were washed three times with PBS and then incubated for 30
353 min at 37 ˚C with donkey anti-rabbit Alexa Fluor 488 antibody (Invitrogen) diluted 1:100 in PBS
354 containing 10 % (v/v) fetal calf serum (FCS). Bacteria were washed three times with PBS before
355 mounting with 20 % Mowiol 4-88 (Calbiochem), 4 mg ml^{-1} *p*-phenylenediamine.

356

357 **Cell infection and N-WASP/F-actin/DNA labeling.** Infection of semi-confluent HeLa cell
358 monolayers with *S. flexneri* was conducted as described (Teh *et al.*, 2012). HeLa cells were
359 grown on sterile round coverslips at the bottom of 24-well trays. Log-phase bacteria were
360 harvested from a 1:50 sub-culture by centrifugation (16000 x g, 2 min, 20 ˚C) and diluted to 3 x
361 10^8 bacteria/mL in Dulbecco's PBS (D-PBS). HeLa Cells were washed with 10 % (v/v) FCS in
362 minimal essential medium (MEM) and then 80 µL of bacteria were added before centrifugation
363 (500 x g, 5 min, 20 ˚C) to assist invasion. After incubation at 37 ˚C with 5 % CO_2 for 1 h, cells
364 were washed three times with D-PBS, and incubated a further 1.5 h with 500 µL of MEM

supplemented with 10 % (v/v) FCS and 40 μg mL^{-1} gentamycin. Cells were then washed three times with D-PBS, fixed for 15 min with 3.7 % (v/v) formaldehyde solution (Sigma) in PBS, and washed twice with PBS. Before staining, cells were incubated with 50 mM NH_4Cl in D-PBS for 10 min, washed with PBS, permeablized with 0.1 % (v/v) Triton X-100 in PBS for 5 min, and washed with PBS. Cells were blocked with 10 % (v/v) FCS in PBS for 20 min, before aspiration and addition of anti-N-WASP diluted at 1:100 in PBS containing 10 % (v/v) FCS for 30 min at 37 ˚C. Cells were washed three times with PBS, and then incubated for 1 h at 37 ˚C with donkey anti-rabbit Alexa Flour 594 antibody (Invitrogen) and Alexa Flour 488 phalloidin (Invitrogen) diluted to 1:100 and 1:200 respectively in PBS containing 10 % (v/v) FCS. After three washes with PBS, DNA was stained with 10 μg mL^{-1} DAPI for 1 min, washed three times with PBS, and mounted for microscopy as described above.

Microscopy. All images of stained bacteria or infected HeLa cells were captured using an Olympus IX-7 Microscope and MetaMorph software (Molecular Devices) with a phase contrast 100 x oil immersion objective and a 1.5 x enlarger. For fluorescence imaging an X-Cite 120Q lamp was used set at high intensity. All bacterial IcsA fluorescence images were acquired with 100 millisecond exposures. All N-WASP fluorescence images were acquired with 500 millisecond exposures. Fluorescence images for background correction were taken for each experiment. IcsA and N-WASP fluorescence images for presentation were recolored using the ICA LUT using ImageJ such that the full intensity spectrum can be easily observed. MetaMorph region measurement tools were used to quantitate fluorescence intensities for individual bacteria. For IcsA quantitation, 50 bacteria were routinely measured for each experiment. For N-WASP recruitment, all bacteria within an infected cell were measured for each experiment.

388

389 **OMP extraction.** OMPs were isolated using differential Sarkosyl treatment (Hobb *et al.*, 2009).
390 5 x 10^{10} log-phase bacteria were collected from a 1:50 sub-culture by centrifugation (3000 x g,
391 20 min, 4 ˚C), resuspended in 15 mL of 10 mM HEPES, pH 7.5, and lysed by sonication. Debris
392 was removed by centrifugation (10000 x g, 10 min, 4 ˚C) and supernatant ultracentrifuged
393 (149000 x g average, 1 h, 4 ˚C). Whole membrane pellets were homogenized in 15 mL of 10
394 mM HEPES, pH 7.5, re-ultracentrifuged (as above), homogenized in 15 mL of 1 % (w/v)
395 Sodium N-lauroylsarcosinate, 10 mM HEPES, pH 7.5, and incubated at 37 ˚C for 30 min with
396 inversion. OMPs were collected by ultracentrifugation (as above), homogenization in 15 mL of
397 10 mM HEPES, pH 7.5, ultracentrifuged a final time (as above), and homogenization in 250 µL
398 of 10 mM HEPES, pH 7.5. OMPs were diluted in 10 x SDS-PAGE loading buffer and heated to
399 100 ˚C for 10 min before analysis.

400

401 ***S. flexneri* pulse-chase proteolysis assay.** The arabinose/glucose (on/off) expression switch of
402 the vector pBAD30 (Guzman *et al.*, 1995) was utilized for controlled expression essentially as
403 described by (Leyton *et al.*, 2014). 1:100 sub-cultures of *S. flexneri* harboring pBADIcsA and
404 derivatives (see Table S1) were grown to log-phase in 100 mL of LB supplemented with glucose
405 before collection by centrifugation (4000 x g, 4 min, 4 ˚C). Bacteria were washed with LB,
406 resuspended in media containing arabinose, and incubated (5 min, 25 ˚C) for IcsA production
407 (pulse). Bacteria were collected (as above), resuspended in 30 mL of media containing glucose,
408 and placed on a 25 ˚C block for 60 min (chase). During the chase, four 1 mL aliquots were taken
409 at times 0 (resuspension), 1, 5, 10, 20, 40, and 60 mins. The protein from the first aliquot was
410 precipitated with 12 % (w/v) final concentration of trichloroacetic acid (TCA) on ice. The second

aliquot was treated with a final concentration of 10 µg mL^{-1} Proteinase K (PK) on ice for 10 min. Proteolysis was then stopped by addition of 4 mM final concentration of phenylmethanesulfonyl fluoride (PMSF) before TCA precipitation (as above). Bacteria from the third aliquot were collected (16000 x g, 1 min, 4 ˚C), treated with 20 µL of chloroform to permeablize the OM (Ames *et al.*, 1984; Wagner *et al.*, 2009), and then PK treated, stopped, and precipitated as described for the second aliquot. All precipitated samples were washed with acetone, dried, and resuspended with 50 µL of SDS-PAGE loading buffer per 1 OD600 unit (measured using the fourth aliquot). Samples were heated to 100 ˚C for 10 min before analysis. The 'relative translocation' was determined by the formula: Relative Translocation=100-100(PK+/PK-) after densitometric analysis using ImageJ.

Database analysis. To generate AT annotation lists, the UniProt knowledgebase (Jain *et al.*, 2009; Magrane *et al.*, 2011) (uniprot.org) was used due to its extensive architecture annotation of protein entries by InterPro (Hunter *et al.*, 2012). Lists which were pruned to exclude fragments. A list of ATs that were PATR exclusive (-PATR) was generated by searching for entries including InterPro cross-references for the AT β-barrel (IPR005546/IPR006315) but not the PATR (IPR013425). Note that IPR013425 incorporates both PFAM (Finn *et al.*, 2014) (PF12951) and TIGRFAM (Haft *et al.*, 2003) (TIGR02601) PATR models. A list of ATs that were PATR inclusive (+PATR) was generated by searching for entries including both the β-barrel and the PATR. This yielded 14518 -PATR entries and 4961 +PATR entries. To remove redundancy, each list was clustered into UniRef100 sequence clusters (Suzek *et al.*, 2007) which groups identical sequences and represents each group with a representative entry. This produced non-redundant -PATR and +PATR lists of 5419 and 2240 representatives respectively. Each

group was further analyzed based on other InterPro annotations, protein length, and PATR copy number.

PATR structural modelling and identification of degenerate PATR. Modeling of the PATR sequence (PF12951 consensus) was achieved using I-TASSER which uses threading and *ab initio* modeling (zhanglab.ccmb.med.umich.edu/I-TASSER) (Roy *et al.*, 2010; Xu *et al.*, 2011). The model had a high C-score of -0.43 (possible range is -5 to 2 were higher values signifies higher confidence in the model), a TM-score of 0.66±0.13 (>0.5 indicates correct topology, <0.17 indicates random similarity. The following templates contributed to the model: Hap (PDB 3SYJ), Ag43 (PDB 4KH3), WlbB (PDB 3MQG), PCSK9 (PDB 2QTW), and KalataB1 (PDB 1JJZ). TM-Align was used to identify degenerate PATR from solved passenger structures (zhanglab.ccmb.med.umich.edu/TM-align) (Zhang *et al.*, 2005). SCOP/CATH protein folds are shared when TM-score > 0.5. All alignment sites are detailed in Table S2.

ACKNOWLEDGMENTS

This work is supported by a Program Grant (No.: 565526) from the National Health and Medical Research Council (NHMRC) of Australia. MTD is the recipient of a Doctor of Philosophy scholarship from the University of Adelaide. We also thank the Research Centre for Infectious Diseases (RCID) for support during this work.

We greatly appreciate Alistair Standish for critical reading of the manuscript. We also appreciate the donations of antisera from both Carol Gross and Thomas Silhavy, and for bioinformatics advice from Dan Kortschak.

REFERENCES

Ames, G.F., Prody, C. and Kustu, S. (1984) Simple, rapid, and quantitative release of periplasmic proteins by chloroform. *J Bacteriol* **160**: 1181-1183.

Bernardini, M.L., Mounier, J., d'Hauteville, H., Coquis-Rondon, M. and Sansonetti, P.J. (1989) Identification of *icsA*, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proc Natl Acad Sci U S A* **86**: 3867-3871.

Besingi, R.N., Chaney, J.L. and Clark, P.L. (2013) An alternative outer membrane secretion mechanism for an autotransporter protein lacking a C-terminal stable core. *Mol Microbiol* **90**: 1028-1045.

Braselmann, E. and Clark, P.L. (2012) Autotransporters: The cellular environment reshapes a folding mechanism to promote protein transport. *The journal of physical chemistry letters* **3**: 1063-1071.

Celik, N., Webb, C.T., Leyton, D.L., Holt, K.E., Heinz, E., Gorrell, R., *et al.* (2012) A bioinformatic strategy for the detection, classification and analysis of bacterial autotransporters. *PloS one* **7**: e43245.

Dai, J.J., Wang, S.H., Guerlebeck, D., Laturnus, C., Guenther, S., Shi, Z.Y., *et al.* (2010) Suppression subtractive hybridization identifies an autotransporter adhesin gene of *E. coli*

IMT5155 specifically associated with avian pathogenic *Escherichia coli* (APEC). *BMC Microbiology* 10.1186/1471-2180-10-236.

Daniels, C. and Morona, R. (1999) Analysis of *Shigella flexneri* Wzz (Rol) function by mutagenesis and cross-linking: Wzz is able to oligomerize. *Mol Microbiol* **34**: 181-194.

Dautin, N. and Bernstein, H.D. (2007) Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu Rev Microbiol* **61**: 89-112.

Drobnak, I., Braselmann, E., Chaney, J.L., Leyton, D.L., Bernstein, H.D., Lithgow, T.*, et al.* (2014) Of linkers and autochaperones: an unambiguous nomenclature to identify common and uncommon themes for autotransporter secretion. *Mol Microbiol* 10.1111/mmi.12838.

Egile, C., Loisel, T.P., Laurent, V., Li, R., Pantaloni, D., Sansonetti, P.J. and Carlier, M.F. (1999) Activation of the CDC42 effector N-WASP by the *Shigella flexneri* IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. *J Cell Biol* **146**: 1319-1332.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R.*, et al.* (2014) Pfam: the protein families database. *Nucleic Acids Research* **42**: D222-D230.

Goldberg, M.B. and Theriot, J.A. (1995) *Shigella flexneri* surface protein IcsA is sufficient to direct actin-based motility. *Proc Natl Acad Sci U S A* **92**: 6572-6576.

Grijpstra, J., Arenas, J., Rutten, L. and Tommassen, J. (2013) Autotransporter secretion: varying on a theme. *Res Microbiol* **164**: 562-582.

Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.

Haft, D.H., Selengut, J.D. and White, O. (2003) The TIGRFAMs database of protein families. *Nucleic Acids Research* **31**: 371-373.

Henderson, I.R. and Lam, A.C. (2001) Polymorphic proteins of *Chlamydia spp*.-autotransporters beyond the Proteobacteria. *Trends Microbiol* **9**: 573-578.

Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C. and Ala'Aldeen, D. (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* **68**: 692-744.

Heras, B., Totsika, M., Peters, K.M., Paxman, J.J., Gee, C.L., Jarrott, R.J., *et al.* (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.

524 Hobb, R.I., Fields, J.A., Burns, C.M. and Thompson, S.A. (2009) Evaluation of procedures for
525 outer membrane isolation from *Campylobacter jejuni*. *Microbiology* **155**: 979-988.
526
527 Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., *et al.* (2012)
528 InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids*
529 *Res* **40**: D306-312.
530
531 Ieva, R. and Bernstein, H.D. (2009) Interaction of an autotransporter passenger domain with
532 BamA during its translocation across the bacterial outer membrane. *Proc Natl Acad Sci U S A*
533 **106**: 19120-19125.
534
535 Ieva, R., Tian, P., Peterson, J.H. and Bernstein, H.D. (2011) Sequential and spatially restricted
536 interactions of assembly factors with an autotransporter beta domain. *Proc Natl Acad Sci U S A*
537 **108**: E383-391.
538
539 Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B.E., *et al.* (2009) Infrastructure
540 for the life sciences: design and implementation of the UniProt website. *BMC bioinformatics* **10**:
541 136.
542
543 Jain, S. and Goldberg, M.B. (2007) Requirement for YaeT in the outer membrane assembly of
544 autotransporter proteins. *J Bacteriol* **189**: 5393-5398.
545

Jong, W.S., ten Hagen-Jongman, C.M., den Blaauwen, T., Slotboom, D.J., Tame, J.R., Wickstrom, D., *et al.* (2007) Limited tolerance towards folded elements during secretion of the autotransporter Hbp. *Mol Microbiol* **63**: 1524-1536.

Junker, M., Besingi, R.N. and Clark, P.L. (2009) Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Mol Microbiol* **71**: 1323-1332.

Junker, M., Schuster, C.C., McDonnell, A.V., Sorg, K.A., Finn, M.C., Berger, B. and Clark, P.L. (2006) Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci U S A* **103**: 4918-4923.

Kajava, A.V. and Steven, A.C. (2006) The turn of the screw: variations of the abundant beta-solenoid motif in passenger domains of Type V secretory proteins. *J Struct Biol* **155**: 306-315.

Kang'ethe, W. and Bernstein, H.D. (2013a) Charge-dependent secretion of an intrinsically disordered protein via the autotransporter pathway. *Proc Natl Acad Sci U S A* **110**: E4246-4255.

Kang'ethe, W. and Bernstein, H.D. (2013b) Stepwise folding of an autotransporter passenger domain is not essential for its secretion. *J Biol Chem* **288**: 35028-35038.

Kida, Y., Taira, J., Yamamoto, T., Higashimoto, Y. and Kuwano, K. (2013) EprS, an autotransporter protein of *Pseudomonas aeruginosa,* possessing serine protease activity induces inflammatory responses through protease-activated receptors. *Cell Microbiol* **15**: 1168-1181.

Kingsley, R.A., Keestra, A.M., de Zoete, M.R. and Baumler, A.J. (2004) The ShdA adhesin binds to the cationic cradle of the fibronectin 13FnIII repeat module: evidence for molecular mimicry of heparin binding. *Mol Microbiol* **52**: 345-355.

Kingsley, R.A., van Amsterdam, K., Kramer, N. and Baumler, A.J. (2000) The *shdA* gene is restricted to serotypes of *Salmonella enterica* subspecies I and contributes to efficient and prolonged fecal shedding. *Infection and Immunity* **68**: 2720-2727.

Lett, M.C., Sasakawa, C., Okada, N., Sakai, T., Makino, S., Yamada, M.*, et al.* (1989) *virG*, a plasmid-coded virulence gene of *Shigella flexneri*: identification of the VirG protein and determination of the complete coding sequence. *J Bacteriol* **171**: 353-359.

Leyton, D.L., Johnson, M.D., Thapa, R., Huysmans, G.H., Dunstan, R.A., Celik, N.*, et al.* (2014) A mortise-tenon joint in the transmembrane domain modulates autotransporter assembly into bacterial outer membranes. *Nature communications* **5**: 4239.

Leyton, D.L., Rossiter, A.E. and Henderson, I.R. (2012) From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nat Rev Microbiol* **10**: 213-225.

Lugtenberg, B., Meijers, J., Peters, R., van der Hoek, P. and van Alphen, L. (1975) Electrophoretic resolution of the "major outer membrane protein" of *Escherichia coli* K12 into four bands. *FEBS letters* **58**: 254-258.

Magrane, M. and Consortium, U. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database : the journal of biological databases and curation* **2011**: bar009.

Makino, S., Sasakawa, C., Kamata, K., Kurata, T. and Yoshikawa, M. (1986) A genetic determinant required for continuous reinfection of adjacent cells on large plasmid in *S. flexneri* 2a. *Cell* **46**: 551-555.

May, K.L. and Morona, R. (2008) Mutagenesis of the *Shigella flexneri* autotransporter IcsA reveals novel functional regions involved in IcsA biogenesis and recruitment of host neural Wiscott-Aldrich syndrome protein. *J Bacteriol* **190**: 4666-4676.

Morona, R., Daniels, C. and Van Den Bosch, L. (2003a) Genetic modulation of *Shigella flexneri* 2a lipopolysaccharide O antigen modal chain length reveals that it has been optimized for virulence. *Microbiology* **149**: 925-939.

Morona, R. and Van Den Bosch, L. (2003b) Lipopolysaccharide O antigen chains mask IcsA (VirG) in *Shigella flexneri*. *FEMS Microbiol Lett* **221**: 173-180.

Morona, R. and Van Den Bosch, L. (2003c) Multicopy *icsA* is able to suppress the virulence defect caused by the *wzz(SF)* mutation in *Shigella flexneri*. *FEMS Microbiol Lett* **221**: 213-219.

Noinaj, N., Kuszak, A.J., Balusek, C., Gumbart, J.C. and Buchanan, S.K. (2014) Lateral opening and exit pore formation are required for BamA function. *Structure* **22**: 1055-1062.

Noinaj, N., Kuszak, A.J., Gumbart, J.C., Lukacik, P., Chang, H.S., Easley, N.C., *et al.* (2013) Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature* **501**: 385-+.

Pavlova, O., Peterson, J.H., Ieva, R. and Bernstein, H.D. (2013) Mechanistic link between beta barrel assembly and the initiation of autotransporter secretion. *Proc Natl Acad Sci U S A* **110**: E938-947.

Peterson, J.H., Tian, P., Ieva, R., Dautin, N. and Bernstein, H.D. (2010) Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proc Natl Acad Sci U S A* **107**: 17739-17744.

Pina-Pedrero, S., Olvera, A., Perez-Simo, M. and Bensaid, A. (2012) Genomic and antigenic characterization of monomeric autotransporters of *Haemophilus parasuis*: an ongoing process of reductive evolution. *Microbiology* **158**: 436-447.

Purdy, G.E., Fisher, C.R. and Payne, S.M. (2007) IcsA surface presentation in *Shigella flexneri* requires the periplasmic chaperones DegP, Skp, and SurA. *J Bacteriol* **189**: 5566-5573.

Rawlings, N.D., Waller, M., Barrett, A.J. and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42**: D503-509.

Renn, J.P. and Clark, P.L. (2008) A conserved stable core structure in the passenger domain beta-helix of autotransporter virulence proteins. *Biopolymers* **89**: 420-427.

Roman-Hernandez, G., Peterson, J.H. and Bernstein, H.D. (2014) Reconstitution of bacterial autotransporter assembly using purified components. *eLife* **3**: e04234.

Rossiter, A.E., Leyton, D.L., Tveen-Jensen, K., Browning, D.F., Sevastsyanovich, Y., Knowles, T.J.*, et al.* (2011) The essential beta-barrel assembly machinery complex components BamD and BamA are required for autotransporter biogenesis. *J Bacteriol* **193**: 4250-4253.

Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**: 725-738.

Ruiz-Perez, F. and Nataro, J.P. (2014) Bacterial serine proteases secreted by the autotransporter pathway: classification, specificity, and role in virulence. *Cell Mol Life Sci* **71**: 745-770.

Shikata, S., Shimada, K., Kataoka, H., Horinouchi, S. and Beppu, T. (1992) Detection of large COOH-terminal domains processed from the precursor of *Serratia marcescens* serine protease in the outer membrane of *Escherichia coli*. *J Biochem* **111**: 627-632.

Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R. and Wu, C.H. (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**: 1282-1288.

Suzuki, T., Mimuro, H., Suetsugu, S., Miki, H., Takenawa, T. and Sasakawa, C. (2002) Neural Wiskott-Aldrich syndrome protein (N-WASP) is the specific ligand for *Shigella* VirG among the WASP family and determines the host cell type allowing actin-based spreading. *Cell Microbiol* **4**: 223-233.

Teh, M.Y., Tran, E.N. and Morona, R. (2012) Absence of O antigen suppresses *Shigella flexneri* IcsA autochaperone region mutations. *Microbiology* **158**: 2835-2850.

Tran, E.N., Doyle, M.T. and Morona, R. (2013) LPS unmasking of *Shigella flexneri* reveals preferential localisation of tagged outer membrane protease IcsP to septa and new poles. *PloS one* **8**: e70508.

van den Berg, B. (2010) Crystal structure of a full-length autotransporter. *J Mol Biol* **396**: 627-633.

Van den Bosch, L., Manning, P.A. and Morona, R. (1997) Regulation of O-antigen chain length is required for *Shigella flexneri* virulence. *Mol Microbiol* **23**: 765-775.

Van den Bosch, L. and Morona, R. (2003) The actin-based motility defect of a *Shigella flexneri rmlD* rough LPS mutant is not due to loss of IcsA polarity. *Microb Pathog* **35**: 11-18.

van Ulsen, P., van Alphen, L., ten Hove, J., Fransen, F., van der Ley, P. and Tommassen, J. (2003) A Neisserial autotransporter NalP modulating the processing of other autotransporters. *Mol Microbiol* **50**: 1017-1030.

Wagner, J.K., Heindl, J.E., Gray, A.N., Jain, S. and Goldberg, M.B. (2009) Contribution of the periplasmic chaperone Skp to efficient presentation of the autotransporter IcsA on the surface of *Shigella flexneri*. *J Bacteriol* **191**: 815-821.

Xu, D., Zhang, J., Roy, A. and Zhang, Y. (2011) Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. *Proteins* **79 Suppl 10**: 147-160.

Zhang, Y. and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**: 2302-2309.

FIGURE LEGENDS

Figure 1: The passenger of IcsA has a single conserved PATR.

A scaled schematic of the AT IcsA (Q7BCK4) is shown **(A)** indicating the signal sequence ($IcsA^{1-52}$) cleaved at the open arrow, the passenger ($IcsA^{53-758}$), and the β-barrel ($IcsA^{759-1102}$).

704 The black arrow indicates the site of specific low efficiency cleavage by IcsP. The passenger has
705 a single copy of the PATR (IcsA$^{526-557}$, red) shown aligned with the PATR sequence (PF12951
706 consensus) with the PATR Hidden Markov Model in **(B)** (pfam.xfam.org/family/PF12951). The
707 positions of four conserved glycines are outlined (orange) and are completely conserved in the
708 model (yellow).
709
710 **Figure 2: PATR mutations lower surface presentation of IcsA.**
711 **(A)** An anti-IcsA immunoblot of total cell samples from log-phase *S. flexneri* expressing IcsA
712 and IcsA-PATR mutants. **(B)** The same bacteria subjected to anti-IcsA immunofluorescence
713 microscopy (IFM). Representative bacteria are shown in phase (top) and fluorescence (bottom)
714 images (4 x 4 μm). IcsA fluorescence was quantitated in **(C)** for n = 5 (where 50 bacteria were
715 measured for each experiment) and analyzed by ordinary one-way ANOVA (Dunnett's, α =
716 0.05). **(D)** OM protein was also extracted from these bacteria using sarkosyl and analyzed by
717 Coomassie Blue staining and immunoblotting. Coomassie staining shows equivalent loading and
718 enrichment of major OMPs. BamA serves as both a positive control for OMP enrichment and a
719 loading control. SurA, Wzz, and DnaK serve as periplasmic, inner membrane and cytoplasmic
720 controls respectively. Total = total bacterial protein sample of *S. flexneri* expressing IcsA. * =
721 degraded IcsA products. **(E)** To indirectly assess intracellular IcsA surface levels, N-WASP
722 recruitment and F-actin accumulation was also tested in infected HeLa cells by IFM. Overlay
723 images are shown (top) for bacterial nucleoids and eukaryotic nuclei detected with DAPI (blue)
724 and actin labelled with phalloidin (green). N-WASP fluorescence images are shown below (20 x
725 20 μm). N-WASP levels were also quantitated in **(F)** for n = 5 (where all bacteria were measured
726 per infected cell for each experiment) and analyzed by ordinary one-way ANOVA (Dunnett's, α

= 0.05). All experiments were conducted using an IcsA and O-antigen deficient strain of *S. flexneri* (RMA2043) expressing IcsA and IcsA-PATR mutants from P_{IcsA} (see Table S1). All fluorescence images are scaled equally relative to each other. WT = wild type, SEM = standard error of the mean, ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Figure 3: PATR mutations decrease the efficiency of IcsA passenger translocation.

Dynamics of passenger translocation was measured by pulse-chase proteolysis assays on live *S. flexneri* expressing IcsA and IcsA-PATR mutants from an arabinose / glucose P_{BAD} switch (see Table S1). 60 minute chase time-courses are shown where bacteria were treated with Proteinase K (PK+), PK and chloroform (PK+/$CHCl_3$+), or not treated (PK-). All experiments were conducted using an IcsA and IcsP deficient strain of *S. flexneri* (RMA4378). **(A)** A mock chase with pBAD30 only. Immunoblot of periplasmic SurA shows proteolysis occurring only after OM permeabilization by $CHCl_3$ treatment. Immunoblot of cytoplasmic DnaK indicates treatments did not result in cytoplasmic protein proteolysis. **(B)** Passenger translocation was chased for IcsA and IcsA-PATR mutants and means quantitated between 5 and 60 min time-points in **(C)**. The means of the relative translocations for IcsA and IcsA-PATR mutants (time-point independent) are shown in **(D)** and analyzed by repeated measures ANOVA (Dunnett's, $\alpha = 0.05$). SEM = standard error of the mean, n = 2, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Figure 4: The importance of the PATR within the AT family.

(A) Alignment of the PATR sites of IcsA and five other ATs; EprS (Q9HY75), ShdA (Q9XCJ4), subtilisin-type serine protease PrtS (P09489) from *Serratia marcescens* (Shikata *et al.*, 1992),

subtilisin-type serine protease BmaA1 (H6T4K9) of *Haemophilus parasuis* (Pina-Pedrero *et al.*, 2012), and NalP (Q8GKS5). Accessions are UniProtKB. Black arrows indicate glycines investigated in this study. Additional PATR sites for ShdA are shown in Figure S1. **(B)** ATs within the UniProt Knowledgebase were grouped by the presence (+PATR) or absence (-PATR) of detectible PATR and further analyzed by InterPro ID domain annotation combinations. The dependency of domain combination on the presence or absence of the PATR is significant ($p < 0.0001$, chi-square). IPR IDs = PectinLyase/P22-like; 012332, 011050, 012334, Pertactin-like (PL); 004899, 003991, 003992, PAP2; 000326, PbH1 (parallel β-helix); 006626, POMP; 003368, PeptidaseS8; 000209, 022398, 023828, 015500, 023827, 017318, PeptidaseS6; 000710, Peptidase S1; 018114, 001254, Lipase; 017186, 001887, 008265, 013831, Vacuolating Cytotoxin; 003842, 004311. Note, the ordering of the domains does not indicate their position within the primary structure, others = all combinations that were < 2 % represented in both groups, No ID = entries that are yet to be annotated. The minimal overlap between the PATR and PL is shown further in Figure S2. **(C)** Lengths frequency histogram. The mean lengths are significantly different (949.9±5.367 and 1453±15.24 for the -PATR and +PATR groups respectively) as tested by two-tailed t-test ($p < 0.0001$). **(D)** PATR copy number per AT (AT) correlates significantly with length (two-tailed $p < 0.0001$, Pearson). **(E)** I-TASSER generated tertiary structure of the PATR. Orientation is a top-down cross-section from N- to C-terminus. The PATR is a predicted triangular wedge with all glycines (red) clustered at the three apexes. To find degenerate PATR this model was spatially aligned to all the solved AT passenger structures using TM-align. Identified degenerate PATR are aligned in **(F)** with the highest scoring site from Ag43 (Q8CVR0) shown below. Spatially conserved glycines (red) between the

772 PATR (blue) and Ag43 (grey) are indicated. For full lists and structure alignments see Table S2
773 and Figure S2.

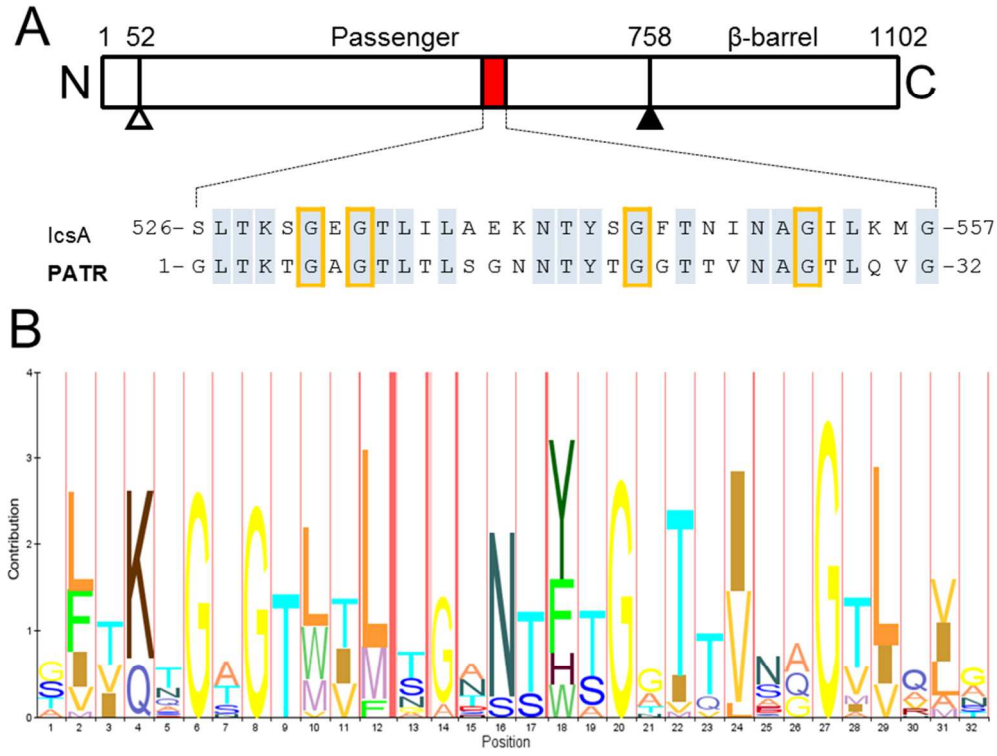

Figure 1: The passenger of IcsA has a single conserved PATR.

A scaled schematic of the AT IcsA (Q7BCK4) is shown (A) indicating the signal sequence (IcsA1-52) cleaved at the open arrow, the passenger (IcsA53-758), and the β-barrel (IcsA759-1102). The black arrow indicates the site of specific low efficiency cleavage by IcsP. The passenger has a single copy of the PATR (IcsA526-557, red) shown aligned with the PATR sequence (PF12951 consensus) with the PATR Hidden Markov Model in (B) (pfam.xfam.org/family/PF12951). The positions of four conserved glycines are outlined (orange) and are completely conserved in the model (yellow).

79x60mm (300 x 300 DPI)

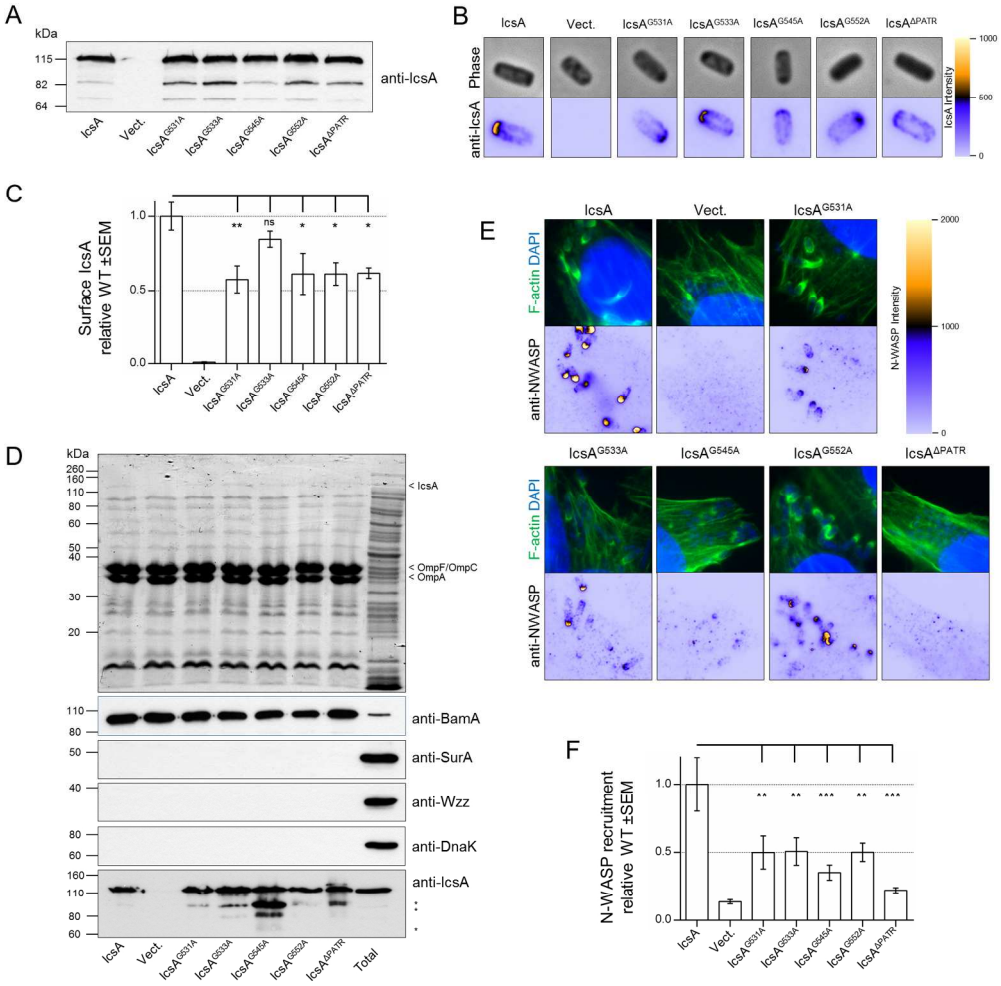

Figure 2 - PATR mutations lower surface presentation of IcsA.

(A) An anti-IcsA immunoblot of total cell samples from log-phase S. flexneri expressing IcsA and IcsA-PATR mutants. (B) The same bacteria subjected to anti-IcsA immunofluorescence microscopy (IFM). Representative bacteria are shown in phase (top) and fluorescence (bottom) images (4 x 4 µm). IcsA fluorescence was quantitated in (C) for n = 5 (where 50 bacteria were measured for each experiment) and analyzed by ordinary one-way ANOVA (Dunnett's, α = 0.05). (D) OM protein was also extracted from these bacteria using sarkosyl and analyzed by Coomassie Blue staining and immunoblotting. Coomassie staining shows equivalent loading and enrichment of major OMPs. BamA serves as both a positive control for OMP enrichment and a loading control. SurA, Wzz, and DnaK serve as periplasmic, inner membrane and cytoplasmic controls respectively. Total = total bacterial protein sample of *S. flexneri* expressing IcsA. * = degraded IcsA products. (E) To indirectly assess intracellular IcsA surface levels, N-WASP recruitment and F-actin accumulation was also tested in infected HeLa cells by IFM. Overlay images are shown (top) for bacterial nucleoids and eukaryotic nuclei detected with DAPI (blue) and actin labelled with phalloidin (green). N-WASP fluorescence images are shown below (20 x 20 µm). N-WASP levels were also quantitated in (F) for n = 5 (where all bacteria were measured per infected cell for each experiment) and analyzed by ordinary one-way ANOVA (Dunnett's, α = 0.05). All experiments were conducted using an IcsA and O-antigen deficient strain of *S. flexneri* (RMA2043) expressing IcsA and IcsA-PATR mutants from P_{IcsA} (see Table S1). All fluorescence images are scaled equally relative to each other. WT = wild type, SEM = standard error of the mean, ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

168x167mm (300 x 300 DPI)

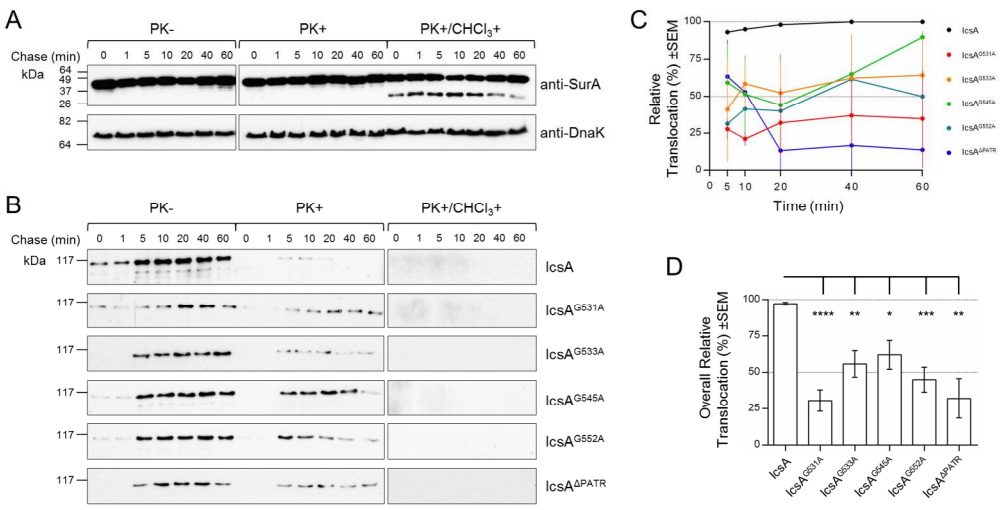

Figure 3: PATR mutations decrease the efficiency of IcsA passenger translocation.

Dynamics of passenger translocation was measured by pulse-chase proteolysis assays on live *S. flexneri* expressing IcsA and IcsA-PATR mutants from an arabinose / glucose P_{BAD} switch (see Table S1). 60 minute chase time-courses are shown where bacteria were treated with Proteinase K (PK+), PK and chloroform (PK+/CHCl3+), or not treated (PK-). All experiments were conducted using an IcsA and IcsP deficient strain of *S. flexneri* (RMA4378). (A) A mock chase with pBAD30 only. Immunoblot of periplasmic SurA shows proteolysis occurring only after OM permeabilization by $CHCl_3$ treatment. Immunoblot of cytoplasmic DnaK indicates treatments did not result in cytoplasmic protein proteolysis. (B) Passenger translocation was chased for IcsA and IcsA-PATR mutants and means quantitated between 5 and 60 min time-points in (C). The means of the relative translocations for IcsA and IcsA-PATR mutants (time-point independent) are shown in (D) and analyzed by repeated measures ANOVA (Dunnett's, α = 0.05). SEM = standard error of the mean, n = 2, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

168x85mm (300 x 300 DPI)

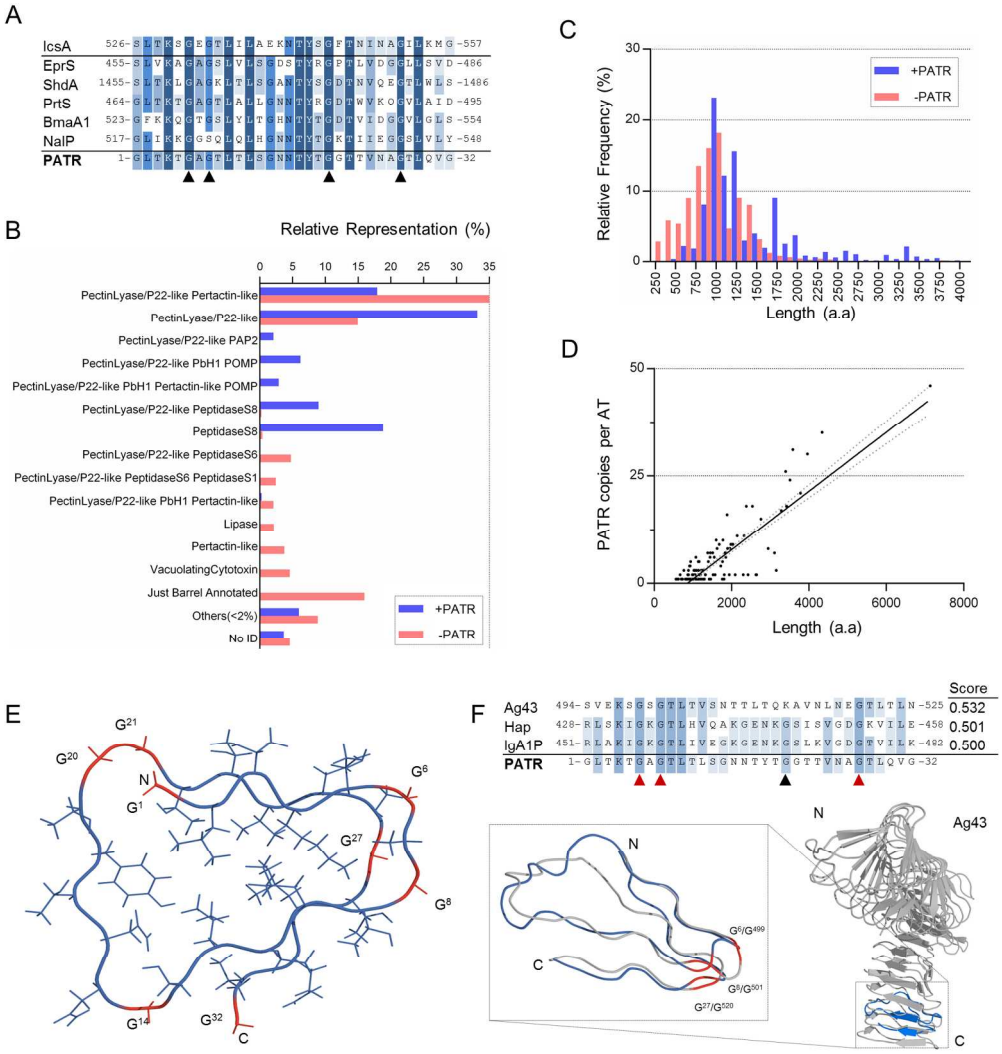

Figure 4: The importance of the PATR within the AT family.

(A) Alignment of the PATR sites of IcsA and five other ATs; EprS (Q9HY75), ShdA (Q9XCJ4), subtilisin-type serine protease PrtS (P09489) from *Serratia marcescens* (Shikata et al., 1992), subtilisin-type serine protease BmaA1 (H6T4K9) of *Haemophilus parasuis* (Pina-Pedrero et al., 2012), and NalP (Q8GKS5). Accessions are UniProtKB. Black arrows indicate glycines investigated in this study. Additional PATR sites for ShdA are shown in Figure S1. (B) ATs within the UniProt Knowledgebase were grouped by the presence (+PATR) or absence (-PATR) of detectible PATR and further analyzed by InterPro ID domain annotation combinations. The dependency of domain combination on the presence or absence of the PATR is significant ($p < 0.0001$, chi-square). IPR IDs = PectinLyase/P22-like; 012332, 011050, 012334, Pertactin-like (PL); 004899, 003991, 003992, PAP2; 000326, PbH1 (parallel β-helix); 006626, POMP; 003368, PeptidaseS8; 000209, 022398, 023828, 015500, 023827, 017318, PeptidaseS6; 000710, Peptidase S1; 018114, 001254, Lipase; 017186, 001887, 008265, 013831, Vacuolating Cytotoxin; 003842, 004311. Note, the ordering of the domains does not indicate their position within the primary structure, others = all combinations that were < 2 % represented in both groups, No ID = entries that are yet to be annotated. The minimal overlap between the PATR and PL is shown further in Figure S2. (C) Lengths frequency histogram. The mean lengths are significantly different (949.9±5.367 and 1453±15.24 for the -PATR and +PATR groups respectively) as tested by two-tailed t-test ($p < 0.0001$). (D) PATR copy number per AT (AT) correlates significantly with

length (two-tailed $p < 0.0001$, Pearson). (E) I-TASSER generated tertiary structure of the PATR. Orientation is a top-down cross-section from N- to C-terminus. The PATR is a predicted triangular wedge with all glycines (red) clustered at the three apexes. To find degenerate PATR this model was spatially aligned to all the solved AT passenger structures using TM-align. Identified degenerate PATR are aligned in (F) with the highest scoring site from Ag43 (Q8CVR0) shown below. Spatially conserved glycines (red) between the PATR (blue) and Ag43 (grey) are indicated. For full lists and structure alignments see Table S2 and Figure S2.

168x180mm (300 x 300 DPI)