



Université de Montréal

**Étude de la biogenèse de l'autotransporteur AIDA-I d'*Escherichia coli***

Par

Marie-Ève Charbonneau

Département de pathologie et microbiologie

Faculté de médecine vétérinaire

Thèse présentée à la Faculté de médecine vétérinaire

en vue de l'obtention du grade de

*philosophiae doctor* (Ph.D.)

en sciences vétérinaires

option microbiologie

Avril 2012

© Marie-Ève Charbonneau, 2012.

Université de Montréal  
Faculté de médecine vétérinaire

Cette thèse intitulée :

Étude de la biogenèse de l'autotransporteur AIDA-I d'*Escherichia coli*

Présentée par :

Marie-Ève Charbonneau

A été évaluée par un jury composé des personnes suivantes :

Josée Harel, présidente-rapporteuse  
Michael Mourez, directeur de recherche  
Christian Baron, membre du jury  
Hervé Le Moual, examinateur externe  
Mario Jacques, représentant du doyen

## Résumé

Les autotransporteurs monomériques, appartenant au système de sécrétion de type V, correspondent à une famille importante de facteurs de virulence bactériens. Plusieurs fonctions, souvent essentielles pour le développement d'une infection ou pour le maintien et la survie des bactéries dans l'organisme hôte, ont été décrites pour cette famille de protéines. Malgré l'importance de ces protéines, notre connaissance de leur biogenèse et de leur mécanisme d'action demeure relativement limitée.

L'autotransporteur AIDA-I, retrouvé chez diverses souches d'*Escherichia coli*, est un autotransporteur multifonctionnel typique impliqué dans l'adhésion et l'invasion cellulaire ainsi que dans la formation de biofilm et d'agrégats bactériens. Les domaines extracellulaires d'autotransporteurs monomériques sont responsables de la fonctionnalité et possèdent pratiquement tous une structure caractéristique d'hélice  $\beta$ . Nous avons mené une étude de mutagenèse aléatoire avec AIDA-I afin de comprendre la base de la multifonctionnalité de cette protéine. Par cette approche, nous avons démontré que les domaines passagers de certains autotransporteurs possèdent une organisation modulaire, ce qui signifie qu'ils sont construits sous la forme de modules fonctionnels.

Les domaines passagers d'autotransporteurs peuvent être clivés et relâchés dans le milieu extracellulaire. Toutefois, malgré la diversité des mécanismes de clivage existants, plusieurs protéines, telles qu'AIDA-I, sont clivées par un mécanisme qui demeure inconnu. En effectuant une renaturation *in vitro* d'AIDA-I, couplée avec une approche de mutagenèse dirigée, nous avons démontré que cette protéine se clive par un mécanisme autocatalytique qui implique deux acides aminés possédant un groupement carboxyle. Ces résultats ont permis la description d'un nouveau mécanisme de clivage pour la famille des autotransporteurs monomériques.

Une des particularités d'AIDA-I est sa glycosylation par une heptosyltransférase spécifique nommée Aah. La glycosylation est un concept plutôt récent chez les bactéries et pour l'instant, très peu de protéines ont été décrites comme glycosylées chez *E. coli*. Nous avons démontré que Aah est le prototype pour une nouvelle famille de glycosyltransférases bactériennes retrouvées chez diverses espèces de protéobactéries. La glycosylation d'AIDA-I est une modification cytoplasmique et post-traductionnelle. De plus, Aah ne



reconnaît pas une séquence primaire, mais plutôt un motif structural. Ces observations sont uniques chez les bactéries et permettent d'élargir nos connaissances sur la glycosylation chez les procaryotes. La glycosylation par Aah est essentielle pour la conformation d'AIDA-I et par conséquent pour sa capacité de permettre l'adhésion. Puisque plusieurs homologues d'Aah sont retrouvés à proximité d'autotransporteurs monomériques putatifs, cette famille de glycosyltransférases pourrait être importante, sinon essentielle, pour la biogenèse et/ou la fonction de nombreux autotransporteurs.

En conclusion, les résultats présentés dans cette thèse apportent de nouvelles informations et permettent une meilleure compréhension de la biogenèse d'une des plus importantes familles de protéines sécrétées chez les bactéries Gram négatif.

### **Mots clés**

Autotransporteur, sécrétion, glycosylation, heptose, clivage autocatalytique, AIDA-I, Aah, adhésion, biofilm, *Escherichia coli*.

## Abstract

Monomeric autotransporters, a family of proteins that use the type V secretion pathway, are important mediators of virulence for many bacterial pathogens. Many functions important for host colonization and survival have been described for these proteins. Despite the recognized importance of this family of proteins, the mechanisms that are required for the biogenesis and functionality of monomeric autotransporters still remain poorly understood.

The *Escherichia coli* adhesin involved in diffuse adherence (AIDA-I) is a classical multifunctional autotransporter protein that mediates bacterial aggregation and biofilm formation, as well as adhesion and invasion of cultured epithelial cells. Extracellular domains of autotransporters are responsible for the protein function and fold into a characteristic  $\beta$ -helical structure. We performed a random mutagenesis of the AIDA-I passenger domain in order to identify regions involved in the various phenotypes associated with the expression of this protein. Our study suggests that the passenger domain of AIDA-I possesses a modular organization, which means that AIDA-I is built with individual functional modules.

Autotransporter passenger domains can be cleaved from the  $\beta$ -domain and released into the extracellular milieu. However, despite the fact that diverse cleavage mechanisms have been previously described, many autotransporters, like AIDA-I, are cleaved by an unknown mechanism. By monitoring the *in vitro* refolding and cleavage following by site-directed mutagenesis, we showed that AIDA-I processing is an autocatalytic event that involves two acidic residues. Our results unveil a new mechanism of auto-processing in the autotransporter family.

AIDA-I is one of the few glycosylated proteins found in *Escherichia coli*. Glycosylation is mediated by a specific heptosyltransferase encoded by the *aah* gene, but little is known about the role of this modification and the mechanism involved. Our findings suggest that Aah represents the prototype of a new large family of bacterial protein *O*-glycosyltransferases that modify various substrates recognized through a structural motif. Furthermore, we showed that glycosylation occurs in the cytoplasm by a cotranslational mechanism. These observations are unique in bacteria and represent a significant advance in our comprehension of prokaryotic glycosylation. We also showed that glycosylation is

required to ensure a normal conformation of AIDA-I and, as a consequence, is necessary for its cell-binding function. The finding that other autotransporters or large adhesin-encoding genes are linked to Aah homologue-encoding genes suggests that glycosylation may be important, if not essential, for the function of these proteins, as for AIDA-I.

In conclusion, the results presented in this thesis bring new information about the autotransporter family and also give new insight into the mechanisms that are important for different aspects of the biogenesis of monomeric autotransporters.

**Key words**

Autotransporter, secretion, glycosylation, heptose, autocatalytic cleavage, AIDA-I, Aah, adhesion, biofilm, *Escherichia coli*.

## Table des matières

<b>Identification du jury</b> .....	<b>ii</b>
<b>Résumé</b> .....	<b>iii</b>
<b>Abstract</b> .....	<b>v</b>
<b>Table des matières</b> .....	<b>vii</b>
<b>Liste des tableaux</b> .....	<b>xiii</b>
<b>Liste des figures</b> .....	<b>xiv</b>
<b>Liste des sigles et abréviations</b> .....	<b>xvii</b>
<b>Remerciements</b> .....	<b>xxi</b>
<b>Introduction</b> .....	<b>1</b>
<b>Recension des écrits</b> .....	<b>3</b>
<b>1. Systèmes de sécrétion bactériens</b> .....	<b>3</b>
1.1. Système de sécrétion Sec.....	3
1.1.1. Le complexe SecYEG.....	4
1.1.2. SecA.....	6
1.1.3. Autres composants du système Sec.....	7
1.1.4. Recrutement des pré-protéines au complexe SecYEG.....	7
1.2. Système de sécrétion TAT.....	10
1.3. Système d'insertion des protéines de membrane externe.....	11
1.4. Autres systèmes de sécrétion.....	14
1.4.1. Système de sécrétion de type I.....	15
1.4.2. Système de sécrétion de type II.....	16
1.4.3. Système de sécrétion de type III.....	17
1.4.4. Système de sécrétion de type IV.....	18
1.4.5. Système de sécrétion de type V.....	19
1.4.6. Système de sécrétion de type VI.....	24

<b>2. La glycosylation chez les bactéries .....</b>	<b>26</b>
2.1. La N-glycosylation .....	27
2.1.1. Système général de N-glycosylation.....	28
2.1.2. N-glycosyltransférases atypiques .....	30
2.2. O-glycosylation .....	30
2.2.1. O-glycosylation des flagelles.....	31
2.2.2. O-glycosylation des pili de type IV .....	33
2.2.3. O-glycosylation des autotransporteurs monomériques de <i>E. coli</i> .....	34
2.2.4. Systèmes généraux de O-glycosylation .....	36
2.2.5. O-glycosylation chez les bactéries Gram positif .....	38
2.3. Applications biotechnologiques .....	40
<b>3. Autotransporteurs monomériques .....</b>	<b>41</b>
3.1. Fonctions et structures des autotransporteurs monomériques .....	42
3.1.1. Le domaine passager.....	43
3.1.2. Le domaine $\beta$ .....	44
3.1.3. Le domaine de jonction.....	45
3.2. Biogenèse des autotransporteurs monomériques .....	45
3.2.1. Transport dans le périplasme .....	46
3.2.2. Transit périplasmique .....	49
3.2.3. Translocation au niveau de la membrane externe.....	51
3.2.4. Clivage.....	54
3.2.5. Autres particularités.....	56
3.3. Utilisation biotechnologiques.....	57
<b>4. AIDA-I.....</b>	<b>58</b>
4.1. Nomenclature .....	58
4.2. Pathogenèse associée à l'expression d'AIDA-I .....	59
4.3. Structure et biogenèse d'AIDA-I .....	61
4.4. Fonctions associées à l'expression d'AIDA-I.....	63

**Article 1 : Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I)**

Implication de l'auteur .....	66
Abstract .....	68
Introduction .....	69
Materials and methods.....	71
Results .....	75
Discussion .....	80
Acknowledgments .....	81
References .....	82
Figure legends .....	87
Table.....	90
Figures.....	91

**Article 2 : Functional Organization of the Autotransporter Adhesin Involved in Diffuse Adherence**

Implication de l'auteur .....	98
Abstract .....	100
Introduction .....	101
Materials and methods.....	102
Results .....	106
Discussion .....	111
Acknowledgments .....	114
References .....	114
Figure legends .....	118
Table.....	121
Figures.....	122

**Article 3 : O-Linked Glycosylation Ensures the Normal Conformation of the Autotransporter Adhesin Involved in Diffuse Adherence**

Implication de l'auteur .....	130
Abstract .....	132
Introduction .....	133

Materials and methods.....	134
Results .....	139
Discussion .....	144
Acknowledgments .....	146
References .....	146
Figure legends .....	150
Table.....	153
Figures .....	155
Supporting information .....	161

**Article 4 : The *Escherichia coli* AIDA-I autotransporter undergoes cytoplasmic glycosylation independently of export**

Implication de l’auteur .....	164
Abstract .....	166
Introduction .....	167
Material and methods .....	168
Results and discussion.....	172
Acknowledgments .....	176
References .....	177
Figure legends .....	179
Figures .....	182

**Article 5 : Autoprocessing of the *Escherichia coli* AIDA-I autotransporter: a new mechanism involving acidic residues in the junction region**

Implication de l’auteur .....	187
Abstract .....	189
Introduction .....	190
Experimental procédures.....	192
Results .....	196
Discussion .....	201
Acknowledgment.....	204
References .....	204
Figure legends .....	207

Table.....	211
Figures.....	212
<b>Article 6 : A structural motif is the recognition site for a new family of bacterial protein O-glycosyltransferases</b>	
Implication de l'auteur .....	221
Summary .....	223
Introduction .....	224
Results .....	226
Discussion .....	232
Experimental procédures.....	234
Acknowledgments .....	238
References .....	238
Figure legends .....	242
Table.....	246
Figures.....	248
Supporting information .....	256
<b>Résultats supplémentaires.....</b>	<b>260</b>
1. Détermination du site catalytique de la glycosyltransférase Aah.....	261
2. Purification de Aah et TibC.....	262
3. Oligomérisation de Aah et TibC .....	264
4. Glycosylation co-translationnelle .....	265
5. Co-purification de composants cytoplasmiques avec AIDA* .....	268
<b>Discussion générale .....</b>	<b>270</b>
1. Glycosylation cytoplasmique .....	270
2. Recrutement d'AIDA-I au complexe SecYEG .....	275
3. Clivage protéolytique d'AIDA-I .....	277
4. Fonctionnalité.....	279



**Conclusion.....282**

**Bibliographie .....285**

**Liste des tableaux****Article 1**

Table I : Primers used in this study.....	90
---	----

**Article 2**

Table I : Primers used in this study.....	121
---	-----

**Article 3**

Table I : Glycosylation of peptides identified in AIDA-I.....	153
---	-----

**Article 5**

Table I : Primers used in this study.....	211
---	-----

**Article 6**

Table I : Predicted glycosyltransferase encoding open reading frames (ORFs) with homology to Aah identified in bacterial genomes. ....	246
Table SI : Primers used in this study.....	256

## Liste des figures

### Recension des écrits

Figure 1 : Sécrétion des protéines bactériennes à travers la membrane plasmique par le système de sécrétion Sec.....	4
Figure 2 : Biogenèse des protéines de membrane chez les bactéries Gram négatif.....	12
Figure 3 : Systèmes de sécrétion retrouvés chez les bactéries Gram négatif.....	14
Figure 4 : Système de sécrétion de type V.....	19
Figure 5 : N-glycosylation chez les bactéries Gram négatif.....	27
Figure 6 : O-glycosylation chez les bactéries Gram négatif.....	30
Figure 7 : Organisation structurale des autotransporteurs monomériques.....	41
Figure 8 : Biogenèse des autotransporteurs monomériques.....	45
Figure 9 : Organisation structurale d'AIDA-I.....	60

### Article 1

Figure 1 : Structural organization of AIDA-I.....	91
Figure 2 : Outer membrane cleavage of AIDA-I is independent of its level of expression.....	92
Figure 3 : Cosolubilization and copurification of AIDA <sub>c</sub> and mature AIDA-I.....	93
Figure 4 : Thermal denaturation of AIDA-I monitored by far-UV CD.....	94
Figure 5 : Effect of modification of the outer membrane cleavage site of AIDA-I.....	95
Figure 6 : Processing of AIDA-I in a protease-deficient strain.....	96
Figure 7 : Effect of the lack of AIDA-I outer membrane cleavage on functionality.....	97

### Article 2

Figure 1 : Effect of insertions on the expression of AIDA-I.....	122
Figure 2 : Effect of insertion mutants on the functionality of AIDA-I.....	123
Figure 3 : Schematic diagram of the insertion mutations in AIDA-I and the related phenotypes.....	124
Figure 4 : Deletion of putative cell-binding domain in mature AIDA-I.....	125
Figure 5 : Function of AIDA-I bearing the $\Delta N$ délétion.....	126
Figure 6 : Cell-binding properties of glycosylated (Glyc.) and unglycosylated (Unglyc.) GST-AIDAN.....	127

Figure 7 : Immunofluorescence detection of purified unglycosylated (Unglyc.) and glycosylated (Glyc.) GST-AIDAN on the surface of epithelial cells .....	128
Figure 8 : Locations of the five-amino-acid insertions in AIDA-I .....	129

### Article 3

Figure 1 : Localization of heptose modifications.....	155
Figure 2 : Effect of glycosylation on the function of AIDA-I .....	156
Figure 3 : Effect of glycosylation on the expression level and on mild heat extraction of AIDA-I.....	157
Figure 4 : Effect of glycosylation on the resistance to protéases.....	158
Figure 5 : Induction of an extracytoplasmic stress by the nonglycosylated AIDA-I.....	159
Figure 6 : Binding of unglycosylated mature AIDA-I to cultured epithelial cells.....	160
Figure S1: Modification of a threonine residue in a peptide obtained by treatment of AIDA-I with chymotrypsin.....	161
Figure S2: DegP and OmpT are not involved in the degradation of unglycosylated AIDA-I.....	162

### Article 4

Figure 1 : Schematic representation of the signal sequence mutants of AIDA-I.....	182
Figure 2 : Subcellular location of the Aah glycosyltransferase.....	183
Figure 3 : Expression and location of the $\Delta$ SS mutant of AIDA-I .....	184
Figure 4 : Cytoplasmic glycosylation of AIDA-I .....	185
Figure 5 : The signal sequence N-terminal extension is not essential for glycosylation ...	186

### Article 5

Figure 1 : Cleavage of AIDA-I* <i>in vitro</i> .....	212
Figure 2 : Protease sensitivity of AIDA-I* during refolding .....	213
Figure 3 : Effect of pH on the <i>in vitro</i> folding of AIDA-I* .....	214
Figure 4 : Mapping of the region involved in AIDA-I cleavage.....	215
Figure 5 : Mutational analysis of the region involved in AIDA-I cleavage.....	216
Figure 6 : Characterization of the D878N protein .....	217
Figure 7 : Renaturation and folding kinetics of D878N* mutant.....	218
Figure 8 : Intramolecular cleavage of AIDA-I.....	219

Figure 9 : Structural model and alignment of the cleavage site of AIDA-I.....	220
---	-----

## Article 6

Figure 1 : Phylogenetic tree of Aah homologues.....	248
Figure 2 : Glycosylation in various bacterial species.....	249
Figure 3 : Glycosylation of a <i>Cronobacter sakazakii</i> protein by Aah .....	250
Figure 4 : Identification of a short sequence sufficient for glycosylation by Aah .....	251
Figure 5 : Glycosylation of the GST–Repeat protein .....	252
Figure 6 : Alanine-scanning of the 19-amino-acid sequence.....	253
Figure 7 : Deletion mutagenesis of the GST–Repeat protein .....	254
Figure 8 : Glycosylation of the <i>B. pertussis</i> pertactin autotransporter protein. ....	255
Figure S1: Alignment of 20 Aah homologs .....	258

## Résultats supplémentaires

Figure 1 : Modèle structural de Aah représentant les résidus importants pour la fonctionnalité .....	260
Figure 2 : Purification et analyse de AahH par CD .....	263
Figure 3 : Purification et analyse de TibCH par CD.....	263
Figure 4 : Oligomérisation de AahH et TibCH.....	264
Figure 5 : Glycosylation <i>in vitro</i> .....	266
Figure 6 : Isolation d’intermédiaires de traduction glycosylés .....	267
Figure 7 : Co-purification de composants cytoplasmiques avec AIDA*.....	268

## Discussion

Figure 1 : Organisation modulaire des SAATs.....	280
--	-----

## Liste des sigles et des abréviations

### Abréviations

Aah	Autotransporter adhesin heptosyltransferase
ABC	ATP-binding cassette
ADN	Acide désoxyribonucléique
ADP	Adénosine diphosphate
ADP-2F-heptose	ADP- <sub>L</sub> -glycero- $\beta$ - <sub>D</sub> -gluco-2-deoxy-3-fluoro-heptose
Ag43	Antigène 43
AGG	Agrégation
AIDA-I	Adhesin involved in diffuse adhérence
Ala, A	Alanine
APEC	Avian pathogenic <i>Escherichia coli</i>
Asn, N	Asparagine
Asp, D	Aspartate
ARN	Acide ribonucléique
ARNt	Acide ribonucléique de transfert
ATP	Adénosine triphosphate
Bam	$\beta$ -barrel assembly machinery
CD	Dichroïsme circulaire
Cys, C	Cystéine
DAEC	Diffuse-adhering <i>Escherichia coli</i>
DO	Densité optique
EPEC	Enteropathogenic <i>Escherichia coli</i>
ESPR	Extended signal peptide region
ETEC	Enterotoxigenic <i>Escherichia coli</i>
Ffh	Fifty-four homologue
Fos	Free oligosaccharides
FPM	Force proton motrice
G/C	Guanine/cytosine
Gln, Q	Glutamine
Glu, E	Glutamate
Gly, G	Glycine

GPS	systeme g6n6ral de s6cr6tion
GTP	Guanosine diphosphate
H	Histidine
Hbp	H6moglobine prot6ase
Hcp	Haemolysin co-regulated protein
HRP	Horseradish peroxydase
IgAP	Immunoglobuline A prot6ase
Ile, I	Isoleucine
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
Kdo	Acide 3-deoxy-D-manno-oct-2-ulosonique
Leu, L	Leucine
Lgf	<i>Bacteroides fragilis</i> glycosylation locus
Lol	Localisation of lipoproteins
LOS	Lipooligosaccharides
LPS	Lipopolysaccharides
Lys, K	Lysine
Met, M	M6thionine
MRE	6llipticit6 moyenne r6siduelle
MTB	Main terminal branch
MW	Molecular weight (poids mol6culaire)
OPG	Osmoregulated periplasmic glucans
Pgl	Protein glycosylation
PlpD	Patatin-like protein
PMF	Prot6ine de fusion membranaire
POTRA	Polypeptide transport associated domains
PPIase	Peptidyl-prolyl isom6rase
Pn	Pectactin
Pro, P	Proline
SAAT	Self-associating autotransporter
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser, S	S6rine
SPATE	S6rine prot6ase autotransporter of <i>Enterobacteriaceae</i>
SRP	Signal recognition particule

SS	Séquence signal
SST1	Système de sécrétion de type I
SST2	Système de sécrétion de type II
SST3	Système de sécrétion de type III
SST4	Système de sécrétion de type IV
SST5	Système de sécrétion de type V
SST6	Système de sécrétion de type VI
STEC	Shiga toxin-producing <i>Escherichia coli</i>
TAM	Translocation and assembly module
TAT	Twin-arginine translocation
T <sub>m</sub>	Température de fusion
TF	Trigger factor
Thr, T	Thréonine
TibA	Enterotoxigenic invasion locus B
TorA	TMAO réductase
TPS	Two partners secretion system
Und-PP	Undecaprenyl pyrophosphate
UPEC	Uropathogenic <i>Escherichia coli</i>
Val, V	Valine
VgrG	Valine-glycine repeat protein
WT	Wild-type (sauvage)

### Symboles

$\alpha$	Alpha
$\beta$	Beta
$\gamma$	Gamma
$\varepsilon$	Epsilon
$\Delta$	Delta

### Unités de mesure

Å	angström
mAu	unité arbitraire
°C	Degré Celsius



cm <sup>2</sup>	centimètre carré
dmol	décimole
g	accélération
Kb	kilo base
kDa	kilo dalton
ml	millilitre
mM	milli molaire
nm	nanomètre
μg	microgramme
μM	micro molaire

## Remerciements

Tout d'abord, je tiens à remercier sincèrement mon directeur de recherche, le docteur Michael Mourez, pour les nombreuses années passées dans son laboratoire. Je le remercie de m'avoir accueillie au tout début, lors d'un premier stage en tant qu'étudiante au baccalauréat et de m'avoir donné le goût de poursuivre en recherche et de vouloir en faire mon métier. Je le remercie aussi pour tout le temps qu'il a passé à me former personnellement, autant d'un point de vue théorique que pratique. Merci beaucoup pour votre confiance et pour m'avoir transmis votre passion pour la science. Vous avez été le meilleur mentor que je n'aurais pu avoir et je vous en serais toujours reconnaissante. Je vous dois beaucoup. Merci pour tout Michael.

Je voudrais aussi remercier les membres de mon comité conseil, soit la docteure Josée Harel et le docteur Hervé Le Moual, pour leur disponibilité et leurs conseils depuis le début de mes travaux dans le laboratoire du docteur Michael Mourez.

J'adresse aussi mes remerciements aux membres actuels ou passés du laboratoire du Dr Mourez pour leur conseil et leur support. Plus particulièrement, je voudrais remercier Frédéric Berthiaume pour avoir eu la patience de me montrer comment fonctionne un laboratoire de recherche et d'avoir répondu à mes trop nombreuses questions au cours de ces dernières années. Je voudrais aussi souligner l'aide que m'a fournie Victoria Girard, aide qui nous a d'ailleurs menées à une publication à titre de co-premières auteures. Finalement, je voudrais remercier Jean-Philippe Côté pour les nombreuses discussions que nous avons eues ainsi que pour son aide précieuse, autant pour son opinion sur mes résultats que pour les nombreux services qu'il m'a rendus au cours des dernières années.

Un gros merci à tous les étudiants, chercheurs, agents de recherche, techniciens et secrétaires du GREMIP, présents ou passés, pour leur temps accordé, conseils, prêts de matériels ou d'équipements et autres. Merci.

Je voudrais aussi remercier tous les organismes subventionnaires qui m'ont apporté un soutien financier inestimable. Parmi ces derniers, figure le Fonds québécois de la recherche sur la nature et les technologies (FQRNT) pour m'avoir accordé une bourse de maîtrise, le

Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) pour ma bourse de doctorat, la Faculté des études supérieures et postdoctorales, la Faculté de médecine vétérinaire, le Centre de recherche en infectiologie porcine (CRIP), le GREMIP et La Cité de biotechnologie de Saint-Hyacinthe.

D'un point de vue plus personnel, je dois mentionner la grande patience et le grand dévouement dont a fait preuve mon conjoint Sébastien Crépin. Il a su me soutenir dans les moments les plus durs et me redonner le moral lors de mes déceptions professionnelles. Merci Sébastien, ta présence à mes côtés a été essentielle pour compléter ce travail de plusieurs années. Je voudrais aussi remercier ma petite Daphnée pour son sourire lorsque je rentre à la maison le soir et pour tous les beaux moments de joie et de détente qu'elle m'a apportés. Mes deux amours, vous m'avez permis de garder une vie équilibrée en plus de m'aider à garder mes priorités à la bonne place. Merci à vous deux.

Finalement, je tiens à remercier ma famille, plus particulièrement ma mère et mon père qui m'ont aidé à finir ma thèse, que ce soit par leur support moral, par le temps qu'ils m'ont accordé (par exemple, en gardant ma fille à de nombreuses occasions) ou par leur aide financière. Merci à vous deux.

## Introduction

La communication entre les bactéries ainsi qu'entre les bactéries et leur environnement, incluant un organisme hôte, est essentielle pour la survie et la prolifération de ces micro-organismes. Pour permettre cette communication, les bactéries exposent à leur surface ou sécrètent dans le milieu environnant une grande variété de protéines (1). Chez les bactéries, ces protéines sécrétées sont initialement synthétisées dans le cytoplasme, ce qui signifie qu'elles doivent traverser des membranes physiologiques (membrane plasmique pour les bactéries Gram positif et membrane interne et externe pour les bactéries Gram négatif) ainsi que l'espace périplasmique dans le cas des bactéries Gram négatif. Afin de permettre la sécrétion de ces protéines à la surface, les bactéries possèdent donc plusieurs systèmes de sécrétion (2).

Parmi ces systèmes de sécrétion, le type V et plus précisément les autotransporteurs monomériques, retrouvé uniquement chez les bactéries Gram négatif, est le plus répandu (3). Une des raisons pouvant expliquer la grande versatilité de ce mode de sécrétion est son exceptionnelle simplicité. En effet, l'appellation autotransporteur provient du fait que toute l'information requise pour la sécrétion d'une protéine est présente sur son seul polypeptide. Ce fait n'exclut toutefois pas que les autotransporteurs utilisent les systèmes généraux bactériens tels que le système de sécrétion général Sec ou le complexe Bam pour l'insertion des protéines de membrane externe ainsi que les chaperonnes cytoplasmiques et périplasmiques générales (4). La plupart des autotransporteurs caractérisés à ce jour jouent un rôle actif dans l'établissement ou le maintien de l'infection bactérienne. D'ailleurs, une grande diversité fonctionnelle a été associée avec l'expression de ces protéines, incluant un rôle dans l'adhésion et l'invasion des cellules épithéliales, l'adhésion aux composants de la matrice extracellulaire, la motilité intracellulaire, la résistance au sérum ainsi que dans diverses activités cytotoxiques (3).

Un des sous-groupes appartenant aux autotransporteurs monomériques est la famille des SAATs (self-associating autotransporters), qui inclut trois protéines retrouvées chez *Escherichia coli*, soit AIDA-I, TibA et Ag43 (5). Ces autotransporters sont retrouvés chez des souches de *E. coli* causant principalement des diarrhées. *E. coli* est considéré comme l'agent infectieux responsable de nombreux cas de diarrhées à travers le monde et est toujours une source majeure de morbidité et mortalité, d'où l'importance de son étude (6). Ces trois autotransporteurs de la famille des SAATs possèdent des caractéristiques

structurales et fonctionnelles communes. Ces protéines sont principalement impliquées dans l'agrégation bactérienne, la formation de biofilm ainsi que l'adhésion cellulaire, trois fonctions importantes pour permettre la survie et la colonisation d'un organisme hôte. Ces protéines partagent aussi une organisation structurale typique des autotransporteurs monomériques, incluant une séquence signal N-terminale permettant le transport dans le périplasme par le système général Sec, un domaine  $\beta$  en C-terminal qui forme un pore dans la membrane externe, ainsi qu'un domaine passager localisé à la surface et correspondant à la région fonctionnelle de la protéine (3). Une autre caractéristique de la famille des SAATs est la glycosylation des domaines passagers par une glycosyltransférase spécifique (7-9). L'étude de la glycosylation protéique chez les bactéries est plutôt récente. Toutefois, on sait maintenant que cette modification peut influencer la fonction de nombreux facteurs de virulence bactériens en plus d'être essentielle pour la physiologie de certaines bactéries (10). À l'opposé, le mécanisme d'action de nombreuses glycosyltransférases bactériennes ainsi que la variété des substrats ciblés par ces dernières demeurent inconnus.

Malgré les nombreuses similarités au sein de la famille des autotransporteurs monomériques, de nombreuses différences persistent et notre compréhension globale de la biogenèse de ces protéines demeure toujours incomplète. Une meilleure connaissance de la fonction et de la biogenèse des autotransporteurs, incluant le mécanisme de glycosylation spécifique aux SAATs, est requise pour bien apprécier leur implication dans la virulence. Le but de ce travail de doctorat était donc de caractériser la fonction et la biogenèse d'un autotransporteur modèle, soit l'adhésine impliquée dans l'adhérence diffuse (AIDA-I) de *E. coli*. Dans cette optique, nous avons étudié, par une approche de mutagenèse aléatoire, la base de la multifonctionnalité de AIDA-I et ainsi découvert une nouvelle fonction de cette protéine, soit l'invasion cellulaire. Nous avons aussi abordé divers aspects de la biogenèse de AIDA-I dont le clivage protéolytique et la glycosylation. Par ces études, nous avons réussi à déterminer un nouveau mécanisme de clivage pour la famille des autotransporteurs en plus de caractériser une nouvelle famille de O-glycosyltransférases bactériennes uniques.

## Recension des écrits

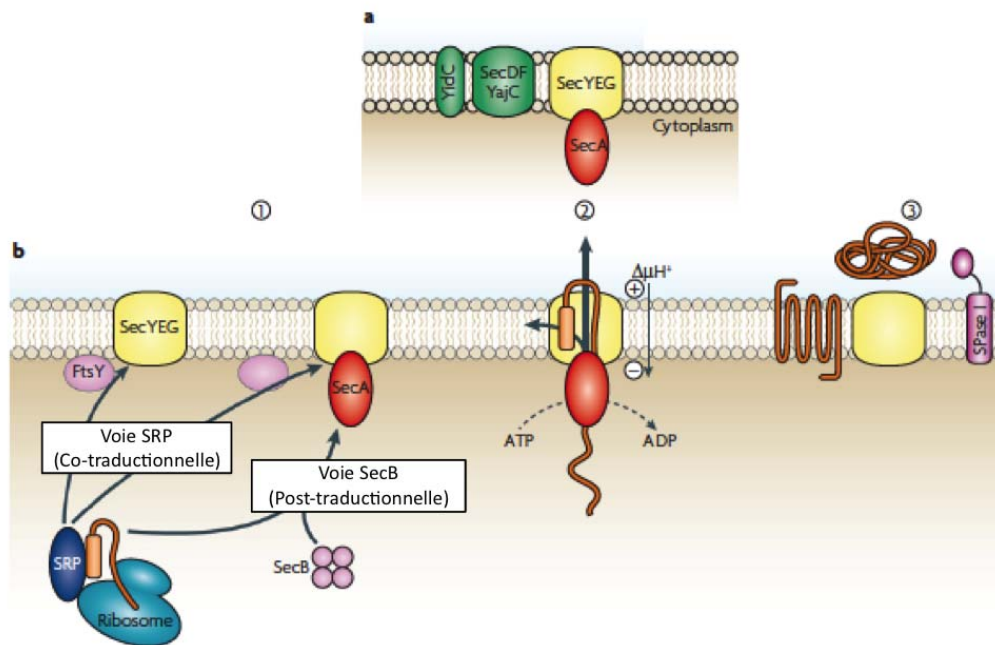
### 1. Systèmes de sécrétion bactériens

La base même de l'interaction entre les bactéries ou entre les bactéries et les cellules d'un organisme hôte repose sur la capacité des bactéries à exposer à leur surface ou à sécréter dans le milieu extracellulaire des molécules effectrices. Pour amener cet arsenal de molécules à la surface, les bactéries doivent posséder des systèmes dédiés au transport de protéines. Ces systèmes doivent permettre aux protéines de traverser la membrane plasmique (membrane interne pour les bactéries Gram négatif) ainsi que le périplasma et la membrane externe des bactéries Gram négatif. Les protéines ainsi sécrétées pourront ensuite interagir avec les cellules hôtes pour permettre, à titre d'exemple, l'adhésion des bactéries, le remodelage du cytosquelette de l'hôte afin de permettre l'invasion ainsi que la modulation de la réponse immunitaire. Toutes ces fonctions sont essentielles pour permettre la survie et la prolifération des bactéries dans l'organisme hôte (1, 11). De plus, les systèmes de sécrétion permettent l'échange d'informations génétiques entre les bactéries, contribuant ainsi à la propagation des gènes de résistances aux antibiotiques. Vu l'importance de ces systèmes de sécrétion pour le pouvoir pathogène des bactéries, leur inhibition est une voie prometteuse pour le développement de nouveaux antibiotiques. Par conséquent, une meilleure connaissance des systèmes de sécrétion bactériens ne peut être que bénéfique. Au cours de la prochaine section, les principaux mécanismes de sécrétion bactériens seront décrits selon la nomenclature établie par Salmond et Reeves (12) et revue par plusieurs autres groupes (3, 13). Cette revue va essentiellement être basée sur les systèmes de sécrétion de protéines retrouvés chez les bactéries Gram négatif.

#### 1.1. Système de sécrétion Sec

La première barrière à laquelle les protéines sécrétées font face est la membrane plasmique. Deux voies de sécrétion majeures existent pour traverser cette membrane, soit le système général de sécrétion Sec et le système de sécrétion TAT (twin-arginine translocation), le plus important étant le système Sec (Figure 1) (14). Ce dernier, présent chez les bactéries Gram positif et Gram négatif ainsi que chez les archaeobactéries, est similaire au système Sec présent dans la membrane du réticulum endoplasmique des cellules eucaryotes (15). Ce

système de sécrétion est d'ailleurs essentiel pour la viabilité dans les trois domaines de la vie. En plus d'être requis pour le transport à travers la membrane plasmique, il est aussi essentiel pour l'intégration de plusieurs types de protéines membranaires. L'importance de ce système est aussi mise en évidence par le fait que plusieurs autres systèmes de sécrétion, qui seront discutés dans les prochaines sections, requièrent un transport Sec-dépendant préalable des substrats dans le périplasma.



**Figure 1 : Sécrétion des protéines bactériennes à travers la membrane plasmique par le système de sécrétion Sec** [figure adaptée de la référence (16)]. **a.** Représentation schématique des différents composants du système général de sécrétion Sec chez les bactéries Gram négatif, incluant le translocon SecYEG (jaune), l'ATPase SecA (rouge), ainsi que les protéines auxiliaires SecDF, YajC et YidC (vert). **b.** Schéma général simplifié représentant le processus de sécrétion. Les pré-protéines (orange) sont dirigées vers le translocon soit par le voie co-traductionnelle SRP-dépendante (bleu) ou par la voie post-traductionnelle SecB dépendante (rose) (étape 1). Les pré-protéines sont sécrétées à travers la membrane interne grâce à l'énergie fournie par l'ATPase SecA et la force proton motrice ( $\Delta\mu\text{H}^+$ ) (étape 2). La signal peptidase permet le clivage des séquences signal du côté périplasmique de la membrane interne (étape 3). Les protéines vont être relâchées dans le périplasma ou intégrées dans la membrane interne.

### 1.1.1. Le complexe SecYEG

Le translocon Sec, c'est-à-dire la partie du système qui forme un pore dans la membrane plasmique et qui permet le passage des polypeptides, est composé de trois protéines, soit

SecY, SecE et SecG (17). Ces trois protéines forment un complexe hétérotrimérique. SecY, une protéine intégrale de membrane forme un complexe stœchiométrique avec la protéine SecE, aussi membranaire, mais de plus petite taille. Ces deux composants sont essentiels pour la viabilité. SecY forme à lui seul le pore servant à la translocation des protéines (18) tandis que SecE est requis principalement pour la stabilité de SecY (19). La troisième composante de ce translocon est la protéine SecG. Cette protéine contient deux segments transmembranaires et interagit directement avec SecY (20). Quoique non essentielle, SecG semble stimuler la sécrétion, et ce, en facilitant l'interaction de SecA (voir plus bas) avec le translocon (21, 22). Toutefois, le mécanisme d'action exact utilisé par SecG reste un sujet de controverse (23).

La première structure publiée du complexe SecYEG provenait de l'archaebactérie *Methanococcus jannaschii* (24). Cette structure représente le complexe sous la forme d'un monomère formant un pore d'environ 5 à 8 Å. Comme une chaîne polypeptidique sous sa forme étendue la plus simple possède une largeur d'environ 4 à 6 Å, le pore formé par le complexe subirait d'important changement de conformation (dilatation ou extension) pour devenir fonctionnel (25, 26). Une étude récente visant à déterminer la taille maximale du canal formé par un monomère SecYEG met en évidence que des molécules avec un diamètre aussi large que 22-24 Å peuvent être efficacement sécrétées (27). Cette dimension de pore serait suffisante pour accommoder des structures secondaires minimales ainsi que des polypeptides liés de façon covalente avec d'autres molécules, telles que des sucres. Cette observation concorde aussi avec la récente structure du complexe Sec61 (orthologue de SecY chez les eucaryotes) en interaction avec un ribosome qui montre clairement qu'un seul monomère est requis pour un système fonctionnel (28).

Toutefois, plusieurs études suggèrent que le complexe SecYEG peut former des oligomères, particulièrement des dimères, dans un contexte physiologiquement relevant, soit dans une membrane en présence des principaux ligands, tels que SecA et les ribosomes (29, 30). Plus spécifiquement, deux modèles pour la formation de dimères, soit le face-à-face (31) ou le dos-à-dos (32), ont été proposés. Le premier modèle suggère que la présence d'un dimère pourrait permettre la formation transitoire d'un pore plus grand et ainsi optimiser le processus de translocation (33). Alternativement, le modèle dos-à-dos suggère qu'une des deux sous-unités du complexe serait active tandis que l'autre serait inactive tout en effectuant des contacts essentiels avec SecA (voir plus bas) (32). Présentement, la controverse autour du fonctionnement de ce complexe demeure d'actualité puisqu'aucun



des modèles existants ne semblent pouvoir rassembler toutes les informations disponibles dans la littérature.

### 1.1.2. SecA

SecA, un composant unique aux eubactéries et aux chloroplastes, est considéré comme le moteur du translocon et fournit l'énergie requise pour le passage des chaînes polypeptidiques à travers la membrane plasmique ainsi que pour l'insertion latérale des protéines membranaires (34). SecA oscille entre le cytoplasme des bactéries, là où elle interagit avec les pré-protéines et la périphérie de la membrane plasmique, où elle s'associe avec une très forte affinité au complexe SecYEG (35). La structure quaternaire de SecA reste un sujet de controverse, malgré la disponibilité de plusieurs structures provenant de différents organismes, soit *Escherichia coli* (36), *Mycobacterium tuberculosis* (37), *Bacillus subtilis* (38) et *Thermus thermophilus* (39). Dans tous ces cas, un dimère a été observé. Par contre, l'orientation des sous-unités dans le dimère varie d'une étude à l'autre. Malgré le fait que ces structures présentent un dimère de SecA, plusieurs études apportent des arguments favorisant un modèle où SecA agirait de façon monomérique (40-42). De plus, une structure récente montre un monomère de SecA en complexe avec SecY, délimitant ainsi l'unité minimale de ce complexe de sécrétion (43). Ces divergences entre les études proviennent principalement du fait que SecA est une molécule dynamique qui, possiblement, oscille entre différentes conformations rendant ainsi l'analyse des résultats plus complexe.

Structuralement, SecA fait partie de la superfamille 2 des protéines contenant un motif DExH/D (44). Cette famille inclus, entre autres, des enzymes capables de modifier les acides nucléiques et des hélicases ATP-dépendantes servant à dérouler l'ADN ou l'ARN. Suite à l'interaction avec une pré-protéine, SecA, lié au complexe SecYEG, va effectuer des cycles d'insertion/désinsertion modulés par l'hydrolyse de l'ATP (45). Chaque cycle va permettre le passage d'environ 30 acides aminés dans le translocon et plusieurs cycles vont être requis pour permettre la sécrétion d'un polypeptide complet (46). Bien que l'hydrolyse de l'ATP est une source d'énergie essentielle pour plusieurs étapes du processus de sécrétion, la force proton motrice (FPM) est aussi utilisée pour alimenter énergétiquement le système (47). La FPM a d'ailleurs été impliquée dans le processus de translocation, pour

stimuler la relâche de l'ADP de SecA suite à l'hydrolyse de l'ATP ainsi que pour l'ouverture de canal formé par SecYEG afin de faciliter la sécrétion.

### 1.1.3. Autres composants du système Sec

Le translocon SecYEG s'associe aussi avec d'autres composants membranaires, dont les protéines SecD et SecF. Ces deux composants forment un complexe et, quoique non essentiel, leur délétion entraîne une inhibition de croissance ainsi qu'un problème sévère de sécrétion (48). La structure de ce complexe a récemment été résolue, permettant ainsi de mieux comprendre le fonctionnement du système Sec (49). Globalement, le modèle suggéré par cette étude propose que ces deux protéines ainsi que la FPM sont requises pour les étapes post initiation du processus de sécrétion, et ce, indépendamment de SecA ou de l'ATP comme source d'énergie. Plus précisément, le domaine périplasmique de SecD interagirait directement avec la chaîne polypeptidique à sa sortie du translocon, ce qui provoquerait un changement de conformation de SecDF, favorisant ainsi le maintien du polypeptide du côté périplasmique. La relâche subséquente du polypeptide par le complexe SecDF et le retour à la conformation initiale seraient couplés avec le mouvement des protons à travers la membrane. Toutefois, d'autres études vont être requises pour confirmer cet attrayant modèle.

Un autre composant associé au système de sécrétion Sec est YidC. Cette protéine semble avoir un rôle unique pour l'intégration des protéines intégrales de membrane (50). Le modèle courant pour le mécanisme d'action de YidC suggère que cette dernière interagit de façon transitoire avec le système Sec, par la formation d'un complexe avec SecDF (51). Il est important de noter que certaines études montrent que YidC peut aussi permettre l'insertion de protéines membranaires, indépendamment du système Sec. Toutefois, peu de substrats ont été identifiés et ces observations impliquent généralement des sous-unités d'importants complexes membranaires comme la  $F_1F_0$ ATP synthase (52).

### 1.1.4. Recrutement des pré-protéines au complexe SecYEG

La reconnaissance des protéines destinées à être sécrétées se fait très rapidement, dès le début du processus de traduction (53). À la sortie du ribosome, les chaînes polypeptidiques peuvent interagir avec la ribonucléoprotéine SRP (signal recognition particule) (54). SRP

est une chaperonne cytoplasmique, formée de la protéine Ffh (fifty-four homologue) et de l'ARN 4.5S (55). Cette interaction entre SRP et le complexe ribosome-chaine naissante provoque une pause ou un ralentissement de la traduction pour ainsi permettre au complexe nouvellement formé de rejoindre le récepteur membranaire de SRP, soit FtsY. Suite à l'hydrolyse du GTP, SRP se dissocie du complexe. FtsY, le ribosome et la chaine naissante vont ensuite interagir avec le translocon SecYEG et ainsi permettre un couplage entre la traduction et la sécrétion (sécrétion co-traductionnelle) (56). Ce mécanisme d'action pour le fonctionnement de la voie SRP est probablement assez simplifié puisque récemment, il a aussi été démontré qu'un complexe membranaire FtsY-SRP pouvait directement recruter un ribosome associé avec une chaine naissante à la membrane plasmique, et ce, indépendamment de la séquence signal (57). Cette étude suggère donc qu'au moins une voie alternative existe pour la sécrétion co-traductionnelle.

Alternativement, les pré-protéines qui ne sont pas reconnues par SRP à la sortie du ribosome peuvent interagir avec le TF (trigger factor) (54). La fixation du TF à une séquence signal va empêcher une liaison subséquente avec SRP et ainsi diriger le polypeptide vers la voie de sécrétion post-traductionnelle. Dans ce cas, la synthèse du polypeptide se fait entièrement dans le cytoplasme avant la sécrétion. Chez plusieurs bactéries (les protéobactéries plus particulièrement), la chaperonne cytoplasmique SecB va venir se fixer sur une région interne de la protéine pour maintenir cette dernière dans un état déplié et compatible avec la machinerie de sécrétion Sec (58). Le complexe SecB-pré-protéine va ensuite être recruté au niveau de la membrane plasmique par une interaction directe avec SecA. La relâche de SecB de ce complexe est liée à la fixation d'une molécule d'ATP sur SecA, ce qui a pour conséquence d'amorcer l'initiation de la sécrétion (59). Il est à noter que SecB ne se fixe pas à toutes les protéines sécrétées de façon post-traductionnelle et peut aussi agir à titre général de chaperonne cytoplasmique (60). D'autres chaperonnes cytoplasmiques, telles que SecA, DnaK et DnaJ, peuvent aussi promouvoir une sécrétion post-traductionnelle (54, 61, 62).

Une question fondamentale par rapport à ce système de sécrétion est la suivante: quel est le signal reconnu par ces différentes voies de recrutement cytoplasmiques? La majorité des protéines qui sont destinées vers les espaces extra-cytoplasmiques sont synthétisées sous forme de pré-protéines, ce qui signifie qu'elles possèdent une séquence signal N-terminale

qui va être clivée suite au passage au travers du translocon par une signal peptidase (63). Une séquence signal typique pour le système Sec fait généralement entre 20 et 30 acides aminés (64). La composition en acides aminés de cette séquence est particulière et peut être divisée en trois sous-domaines : Un domaine N-terminal, nommé domaine n, qui contient au moins un acide aminé chargé positivement, un domaine central h qui est riche en résidus hydrophobes et le domaine polaire c qui contient le motif reconnu pour le clivage par la signal peptidase. Plusieurs des chaperonnes mentionnées précédemment (SRP et le TF principalement) interagissent avec la protéine L23 du ribosome, les plaçant directement au site de sortie du polypeptide. De plus, SecA semble aussi pouvoir reconnaître une séquence signal dès son émergence du ribosome (54). Il est donc généralement accepté que, dès que le polypeptide commence à être exposé, son interaction avec une ou plusieurs de ces chaperonnes cytoplasmiques va guider la pré-protéine vers une des voies menant au système de sécrétion Sec.

Depuis longtemps, il est connu que SRP se lie préférentiellement aux séquences signal hydrophobes (65). Quoique cette hydrophobicité soit nécessaire, elle n'est toutefois pas suffisante pour expliquer la reconnaissance par SRP (66). En effet, d'autres facteurs, tels que la prédisposition des polypeptides à former des structures hélicales ainsi que la présence de résidus à nature basique dans la région N-terminale de la séquence signal, peuvent influencer la liaison de SRP (65, 67). Il est à noter que les substrats bactériens de cette chaperonne sont majoritairement des protéines intégrales de membrane (68). Un motif de reconnaissance a aussi été suggéré pour le TF. Ce dernier correspond à une séquence de 8 acides aminés possédant un nombre élevé de résidus basiques ou aromatiques et possédant une charge nette positive (69). Ce motif revient environ tous les 32 acides aminés dans un polypeptide déplié mais est beaucoup plus rare dans une protéine correctement repliée. Le modèle présentement suggéré est donc que SRP et le TF sont en compétition pour une interaction avec le polypeptide qui émerge du ribosome (70). La présence d'un segment hydrophobe stabiliserait l'interaction avec SRP tout en inhibant la fixation du TF. Toutefois, étant donné la présence d'une multitude de chaperonnes et enzymes à proximité des chaînes naissantes à leur sortie du ribosome, ce modèle paraît bien simplifié.

Une nouvelle avenue d'étude suggère que les propriétés structurales du polypeptides à l'intérieur même de ribosome pourraient servir de signal entraînant le recrutement de différents facteurs au site de sortie du ribosome (71, 72). Une étude décrit d'ailleurs ce type d'interaction pour le translocon Sec61p chez les eucaryotes (73). Quoique que peut de

données sont actuellement disponibles sur ce type de « signal » servant au recrutement de chaperonnes, ce nouveau mode de signalisation intra-ribosomal va certainement prendre de l'ampleur au cours des prochaines années.

## 1.2. Système de sécrétion TAT

Le second système de sécrétion général localisé dans la membrane plasmique est le système TAT. Contrairement au système Sec, qui permet la sécrétion de polypeptides dépliés uniquement, le système TAT transporte à travers la membrane des protéines repliées (74). Chez la plupart des bactéries, seulement un petit nombre de protéines vont utiliser cette voie de sécrétion alors que chez les archaebactéries, ce système est beaucoup plus important (75, 76). La plupart des substrats de TAT sont des protéines de type oxido-réductases qui ont besoin de co-facteurs pour leur action, telles que certaines enzymes impliquées dans la respiration anaérobie (77).

Les substrats destinés au système TAT possèdent une séquence signal différente de celle reconnue par le système Sec (78). À titre comparatif, la région n des séquences signal pour le système de sécrétion TAT est généralement plus longue, la région h moins hydrophobe et la région c qui, riche en acides aminés chargés positivement, sert de motif d'exclusion pour le système Sec. De plus, ces séquences signal ont un motif caractéristique retrouvé entre la région n et h, soit la séquence Z-R-R-x-□-□, où Z est un résidu polaire et □ un résidu hydrophobe. Les deux arginines sont les principales composantes de ce motif.

La sécrétion par le système TAT se fait de façon post-traductionnelle puisque les substrats sont repliés avant leur transport. Cela signifie que ces polypeptides doivent être pris en charge par des chaperonnes cytoplasmiques pour promouvoir un repliement adéquat ainsi que pour éviter d'être dirigés vers le système Sec ou de subir une sécrétion prématurée via le système TAT. Les chaperonnes cytoplasmiques générales qui interagissent avec la séquence signal des ces substrats sont principalement DnaK et SlyD (79). Toutefois, certaines protéines possèdent leur propre chaperonne. C'est le cas pour la TMAO réductase (TorA) de *E. coli* qui possède sa chaperonne spécifique TorD. TorD permet ainsi un repliement efficace de TorA tout en prévenant une dégradation prématurée de la séquence signal (80).

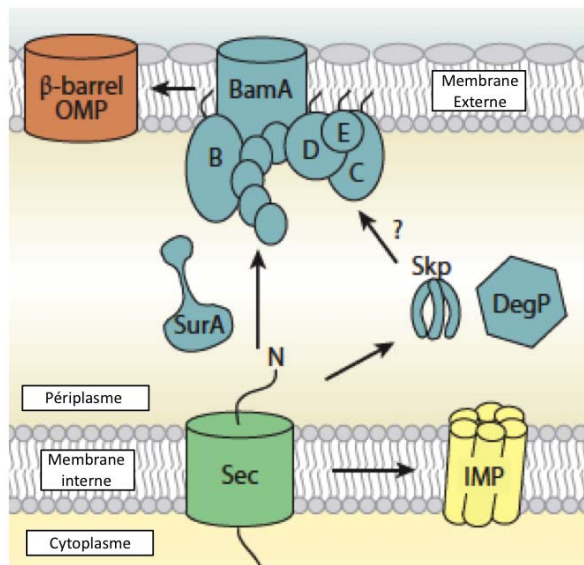
Le translocon du système Tat est composé d'un minimum de deux protéines, TatA et TatC (81). Toutefois, chez les protéobactéries ainsi qu'au niveau des chloroplastes, une troisième

composante, TatB, est requise pour le fonctionnement optimal du système (82). TatB et TatC forment un complexe membranaire chez *E. coli* qui va permettre le recrutement des substrats au translocon (83). C'est la protéine TatC qui va lier directement le motif double arginine de la séquence signal (84). TatA est recruté au complexe TatBC uniquement suite à la fixation d'un substrat, en présence de la source d'énergie requise pour ce système de sécrétion, soit la FPM (84). Cette protéine forme des oligomères de taille importante qui peuvent varier entre 34 kDa et plus de 500 kDa et représente le composant qui forme le pore membranaire (85). Le mécanisme utilisé par ce système de transport consiste à moduler la dimension du pore formé par TatA afin de laisser passer des substrats de grosseur variable tout en préservant l'intégrité membranaire. Ce système de sécrétion est donc bien différent du système Sec et n'est requis, du moins chez les protéobactéries, que pour quelques protéines avec une biogenèse bien particulière.

### 1.3. Système d'insertion des protéines de membrane externe (complexe Bam)

Les protéines transportées à travers la membrane interne des bactéries Gram négatif peuvent ensuite rester dans le périplasma ou se diriger vers la membrane externe pour y être intégrées ou être sécrétées dans le milieu extracellulaire. Jusqu'à récemment, le mécanisme impliqué dans l'insertion des protéines de membrane externe était inconnu. Ce dernier a été découvert chez *Neisseria meningitidis* en 2003 (86). Malgré la découverte plutôt récente, les connaissances sur ce système sont maintenant plutôt nombreuses et plusieurs structures sont aussi disponibles pour différents composants du système Bam ( $\beta$ -barrel assembly machine) (Figure 2).

À la sortie du complexe Sec, les protéines de membrane externe doivent traverser le périplasma tout en restant dans un état de repliement compatible avec leur destination finale. Pour ce faire, ces protéines vont être prises en charge par des chaperonnes périplasmiques. La principale chaperonne impliquée dans le transport de ces protéines est SurA, une PPIase (peptidyl-prolyl-isomérase) (87). Une voie alternative existe et requière les chaperonnes Skp et DegP. Toutefois, cette seconde voie serait moins importante et servirait surtout en cas de stress important. D'autres PPIases, telles que FkpA, PpiA et PpiD pourraient aussi jouer un rôle secondaire dans le transport de ces protéines de membrane externe.



**Figure 2 : Biogenèse des protéines de membrane chez les bactéries Gram négatif** [figure adaptée de la référence (88)]. Les protéines intégrales de membrane sont synthétisées dans le cytoplasme puis transportées dans le périplasm par le système Sec. Les protéines intégrales de membrane interne sont intégrées dans la membrane lors du processus de translocation par le système Sec. Les protéines destinées à la membrane externe vont traverser le périplasm dans un état majoritairement déplié grâce à l'aide de chaperonnes périplasmiques telles que SurA, Skp et DegP. Le repliement et l'insertion dans la membrane externe se fait avec l'aide du complexe Bam (BamA - BamE).

Le composant central du système Bam est la protéine membranaire essentielle BamA (89). On retrouve des homologues de BamA chez toutes les bactéries Gram négatif ainsi que dans les mitochondries et chloroplastes (90). Cette protéine est ancrée dans la membrane externe par un domaine qui forme un tonneau  $\beta$ . La structure du domaine périplasmique de BamA retrouvé chez *E. coli* a été résolue et consiste en cinq domaines POTRA (polypeptide transport associated domains) (91, 92). Chacun de ces domaines se replie individuellement et forment un arrangement composé de deux hélices  $\alpha$  repliées sur un feuillet  $\beta$  comprenant 3 brins. Ces domaines POTRA interagissent avec les substrats destinés à la membrane externe tout en faisant des contacts avec les quatre autres composants du système, soit les lipoprotéines BamB, C, D et E.

Comme pour tout système de sécrétion/insertion de protéines, la question du signal reconnu par la machinerie pour identifier les substrats potentiels est un point déterminant. Dans le cas du complexe Bam, il a été suggéré qu'un motif, qui consiste en une phénylalanine C-terminale ainsi que la présence de résidus hydrophobes aux positions 3, 5, 7 et 9 depuis l'extrémité C-terminale, servirait de signal de reconnaissance (93). D'autres études ont

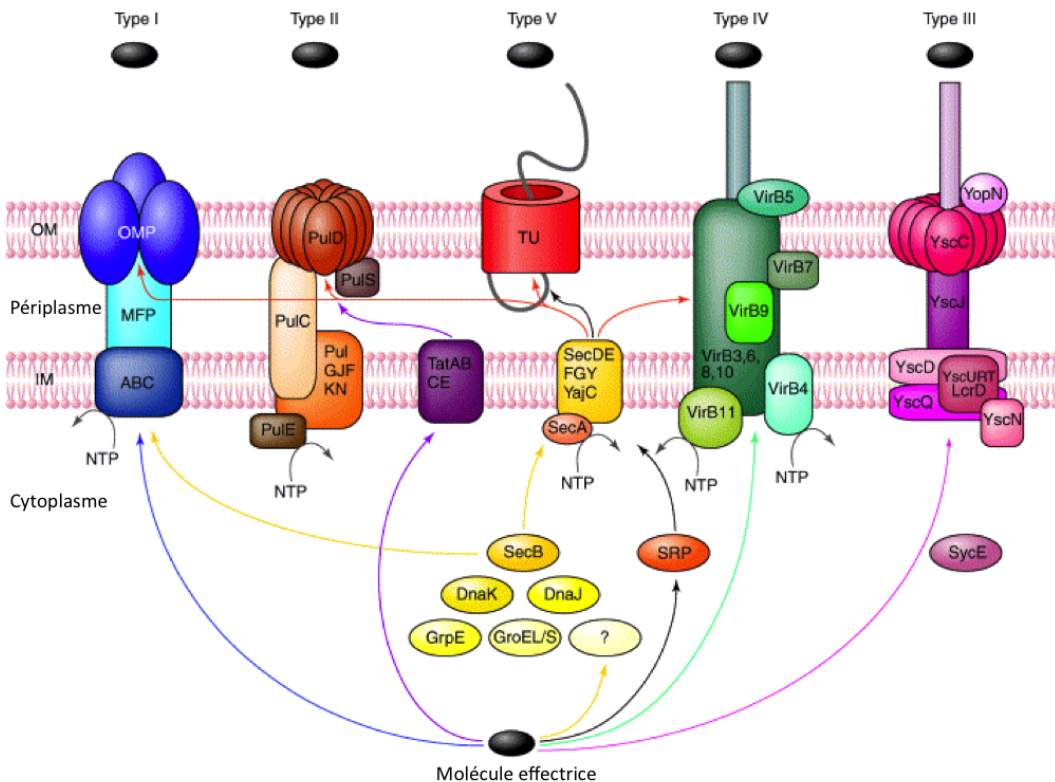
aussi montré, par pontage chimique, que SurA se situait à proximité de BamA et BamB (87). Ceci suggère un modèle attrayant où un polypeptide qui sort du translocon Sec serait pris en charge par la chaperonne SurA et ensuite dirigé vers le complexe Bam, où BamB faciliterait le recrutement de SurA lié à une protéine substrat puis le transfert vers BamA. Toutefois, ce modèle reste à confirmer.

Le rôle des chaperonnes périplasmiques est de garder un polypeptide dans un état principalement déplié. Cela signifie que le complexe Bam est impliqué autant dans le repliement que dans l'insertion des protéines de membrane externe. Basé sur les structures disponibles des domaines POTRA de BamA, il a été suggéré que ces derniers peuvent avoir un rôle dans le repliement des substrats par un mécanisme nommé l'augmentation  $\beta$  (94). Il est à noter que la vaste majorité des protéines intégrales de membrane externe ont une structure riche en brins  $\beta$ . Dans ce modèle, l'interaction de BamA avec son substrat se ferait principalement par la reconnaissance d'une structure secondaire et serait l'élément déclencheur pour l'initiation du repliement (92). Toutefois, comme les domaines POTRA sont périplasmiques, il est logique de penser que d'autres composants du complexe vont être impliqués dans les étapes plus tardives du repliement ainsi que dans le processus même d'insertion dans la membrane externe. Ces étapes sont d'ailleurs les moins bien connues et plusieurs modèles ont été proposés. Le premier modèle suggère que les nouvelles protéines auraient accès à la membrane externe via le centre du pore formé par le tonneau  $\beta$  de BamA. Ceci signifie que le tonneau de BamA devrait se « briser », c'est-à-dire que les liens hydrogènes maintenant ensemble les brins  $\beta$  devraient être temporairement rompu pour permettre la diffusion de la nouvelle protéine dans la membrane. Cette hypothèse est appuyée par une récente étude qui démontre que BamA forme un tonneau  $\beta$  moins stable que la plupart des autres protéines de membrane externe chez *E. coli* (95). Le deuxième mécanisme évoqué propose que différents complexes Bam formeraient un oligomère de grande taille, générant ainsi une cavité centrale hydrophobe permettant l'insertion d'une nouvelle protéine. Ce modèle se base principalement sur l'observation que BamA peut former des oligomères *in vitro* (93). BamCDE, autres composants du système, pourraient aussi être impliqués dans cette étape tardive d'insertion. Malgré le fait que plusieurs structures sont disponibles pour ces composants (96, 97), on possède toujours très peu d'informations sur leur rôle fonctionnel exact et tous ces modèles sont encore au stade hypothétique.



#### 1.4. Autres systèmes de sécrétion

Comme mentionné au début de cette section, les protéines sécrétées sont essentielles pour les interactions entre les bactéries et leur environnement ainsi qu'avec un organisme hôte, et par conséquent, les bactéries possèdent une multitude de système permettant la sécrétion de molécules dans le milieu extracellulaire. Au cours de cette section, quelques uns des systèmes de sécrétion spécifiques aux bactéries Gram négatif seront brièvement discutés (Figure 3). Toutefois, ceci ne correspond pas à une revue exhaustive de tous les systèmes de sécrétion bactériens, puisque les systèmes spécifiques à l'assemblage des pili (98), des curli (99), des lipoprotéines (100), des LPS (101) ainsi que le système de sécrétion de type VII absent chez les bactéries Gram négatif (102), ne seront pas discutés.



**Figure 3 : Systèmes de sécrétion retrouvés chez les bactéries Gram négatif** [figure adaptée de la référence (103)]. Représentation schématique des systèmes de sécrétion bactériens de type I à V. Les connections entrent les différents systèmes sont illustrées. Abréviations : IM, membrane interne; OM, membrane externe; TU, unité de translocation (aussi appelé domaine  $\beta$ ); NTP : nucléotide tri-phosphate; MFP, protéine membranaire de fusion; ABC, ATP-binding cassette exporter; OMP, protéine de membrane externe.

#### 1.4.1. Système de sécrétion de type I

Le système de sécrétion de Type I permet à un substrat d'être sécrété directement dans le milieu extracellulaire à partir du cytoplasme, et ce, sans intermédiaire périplasmique stable. Trois composants majeurs sont requis pour permettre la sécrétion, soit un ABC (ATP-binding cassette exporter) transporteur dans la membrane interne, une protéine membranaire de fusion (PFM) aussi ancrée dans la membrane interne mais possédant un large domaine périplasmique, ainsi qu'une protéine de membrane externe présentant une homologie structurale avec la protéine trimérique TolC (104). Les substrats de ce système de sécrétion varient énormément en terme de fonction, incluant des adhésines, des protéases, des phosphatases, des lipases, des nucléases ainsi que des toxines (105). Le premier substrat identifié a été l'alpha hémolysine (HlyA) sécrétée par des souches de *E. coli* uropathogènes (106). Parmi les protéines sécrétées par ce système, les plus connues sont l'adénylate-cyclase de *Bordetella pertussis* (107) ainsi que la Colicine V de *E. coli* (108). Le signal reconnu par cette machinerie correspond à une séquence C-terminale pouvant contenir environ 50 acides aminés et qui n'est pas clivée lors du processus de sécrétion (109). Toutefois, aucun motif particulier dans cette séquence n'a pu être identifié à ce jour. Étant donné la localisation de ce signal, le système de sécrétion de Type I transporte des polypeptides de façon post-traductionnelle.

Le modèle général pour le mécanisme d'action suggère que c'est l'ABC transporteur qui reconnaît les substrats, dicte la spécificité du système et permet le transport à travers la membrane interne. Toutefois, certaines protéines, telle que la dispersine de *E. coli*, semblent être sécrétées par ce système à partir du périplasme, utilisant le système Sec pour accéder à ce compartiment (110). Les autres composants, soit la protéine adaptatrice PMF ainsi que de la protéine formant le pore dans la membrane externe (TolC), vont être recrutés de façon séquentielle uniquement en présence d'un substrat (105). Il est aussi à noter que la fonction de TolC n'est pas restreinte à la sécrétion de polypeptides en provenance du système de sécrétion de type I, mais est aussi un composant des pompes à efflux bactériennes (111).

#### 1.4.2. Système de sécrétion de type II

Le système de sécrétion de type II a été le premier système de sécrétion spécifique décrit chez les bactéries Gram négatif. Historiquement, ce système a déjà reçu l'appellation de branche terminale principale (MTB) du système général de sécrétion (GPS) (112). Le modèle d'étude pour ce système est celui trouvé chez *Klebsiella oxytoca*, qui permet la sécrétion de la pullulanase PulA (113, 114). Ce système permet la sécrétion de plusieurs types de protéines, telles que des toxines, protéases, cellulases ou lipases, et a été relié à la pathogenèse de certaines bactéries, dont *Legionella pneumophila* (115).

Le système de sécrétion de type II forme un complexe trans-enveloppe qui comporte entre 12 et 15 composants (116). Au niveau de la membrane interne, plusieurs protéines (T2S L, M et F) forment un complexe stable qui va servir de plateforme pour la liaison au composant T2S E, une ATPase cytoplasmique qui fournit l'énergie requise pour la sécrétion. Ce complexe est relié aux composants de la membrane externe par une protéine (nommée T2S C) ancrée au niveau de la membrane plasmique. Les composants de la membrane interne servent aussi d'ancrage pour la formation d'un pseudopilus, structure qui traverse le périplasme et qui présente une homologie structurale avec les pili de type IV (117, 118). Ce pseudopilus est formé par la sous-unité majeure T2S G ainsi que par quatre sous-unités mineures, soit H, I, J et K, où les composants I, J et K forment l'extrémité du pseudopilus. Au niveau de la membrane externe, le composant majeur est la sécrétine (T2S D). Dans une structure obtenue par microscopie électronique, il est montré que la sécrétine de *Vibrio cholerae* forme un oligomère possédant un diamètre interne de 55 Å, soit une dimension suffisante pour permettre le passage de la toxine cholérique complètement repliée (119). L'assemblage de la sécrétine au niveau de la membrane externe requiert la présence d'une pilotine spécifique qui prévient une oligomérisation précoce de la sécrétine (120).

Les substrats du système de sécrétion de type II sont sécrétés par un processus qui requiert deux étapes distinctes (116). Tout d'abord, les protéines cibles sont transportées dans le périplasme par l'intermédiaire des systèmes généraux de sécrétion, soit Sec ou Tat. Dans le périplasme, les substrats complètement repliés vont interagir avec la machinerie de sécrétion et vont être poussés vers la sécrétine par l'action du pseudopilus, qui agit de façon similaire à un piston.

### 1.4.3. Système de sécrétion de type III

Le système de sécrétion de type III est un système de sécrétion bactérien plus complexe que ceux décrits précédemment. Ce dernier est composé d'un minimum de 20 composants structuraux et peut compter en général entre 4 et 20 protéines sécrétées, sans mentionner que certaines de ces protéines requièrent la présence d'une chaperonne cytoplasmique spécifique (121). La beauté de ce système repose sur sa capacité à prendre en charge des molécules effectrices du cytoplasme bactérien pour les introduire directement dans le cytosol de la cellule eucaryote ciblée. Ce mode de livraison des substrats est particulièrement efficace et limite les risques de dégradation prématurée des molécules effectrices par un passage dans le milieu extracellulaire.

Ce système de sécrétion est retrouvé chez plusieurs espèces bactériennes pathogènes, telles que *Pseudomonas*, *Yersinia*, *Shigella*, *Salmonella* et *E. coli*. En fait, il existe présentement sept familles de systèmes de sécrétion de type III, classification principalement basée sur une homologie entre les composants structuraux du système (121). Toutefois, le profil de protéines effectrices varie énormément entre les différentes espèces bactériennes, reflétant une différence dans leurs modes de vie respectifs (122). Par contre, ces effecteurs vont souvent cibler des protéines eucaryotes impliquées dans le réarrangement du cytosquelette, le maintien de l'intégrité membranaire ou l'induction de l'apoptose (123).

Globalement, cette machinerie de sécrétion peut se diviser en trois complexes distincts, soit le corps basal, la « seringue » et le translocon (124). Le corps basal correspond à la partie de la machinerie qui traverse l'enveloppe bactérienne en formant une structure en forme d'anneau au niveau de la membrane interne et externe. Ce complexe est connecté à une ATPase, protéine de membrane interne qui fait face au cytoplasme et qui fournit l'énergie au complexe. Le corps basal est associé à la seringue, qui s'étend au delà de la membrane externe dans le milieu extracellulaire. Cette seringue agirait comme canal permettant le passage des effecteurs. Le diamètre interne de la seringue fait entre 2 et 2.5 nm, permettant le passage de protéines dépliées ou minimalement repliées uniquement (125). Le translocon ne s'assemble que lorsque qu'un contact préalable entre la bactérie et la cellule eucaryote cible a été établi (126). Les composants du translocon sont sécrétés via le corps basal et la seringue pour atteindre la surface de la cellule cible. Leur fonction principale est de former un pore dans la membrane de la cellule eucaryote permettant ainsi le passage subséquent

des protéines effectrices. Évidemment, l'assemblage de cette machinerie ainsi que la sécrétion des effecteurs sont des processus soumis à une régulation complexe (127). Il est à noter que cette machinerie de sécrétion est très similaire à celle permettant la formation du flagelle, et que ces deux systèmes proviendraient d'un ancêtre commun (128).

#### 1.4.4. Système de sécrétion de type IV

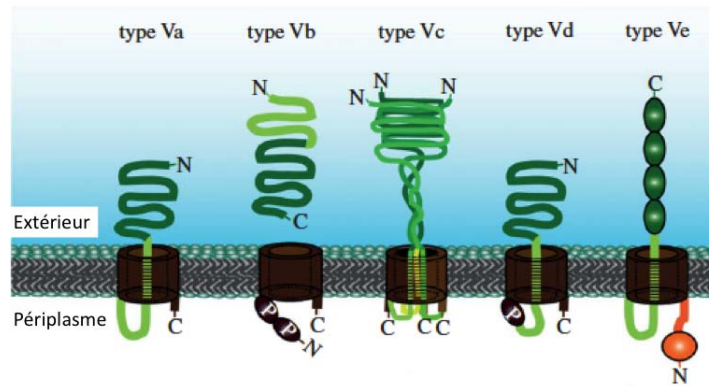
Le système de sécrétion de type IV (SST4) est particulier puisque, en plus de promouvoir la sécrétion de protéines effectrices, il permet aussi le transfert d'ADN. Ce transfert d'ADN entre souches/espèces bactériennes est aussi reconnu comme une cause possible de l'émergence de résistance multiple aux antibiotiques (129). Trois types de SST4 ont été décrits (130). Le premier correspond aux systèmes de conjugaison, qui permettent le transfert d'ADN d'une cellule à une autre de façon contact-dépendant. Ce type de SST4 est le plus répandu et se retrouve aussi chez les bactéries Gram positif et certaines espèces d'archaebactéries. Le deuxième type permet l'acquisition ou la relâche d'ADN directement à partir du milieu environnant. Ce système est toutefois retrouvé uniquement chez *Helicobacter pylori* (système Com, import d'ADN) et *Neisseria gonorrhoeae* (système GGI, relâche d'ADN) (131, 132). Le dernier type de SST4 sert à la sécrétion d'effecteurs bactériens et sera discuté plus en détail dans cette section.

Ce système de sécrétion est un élément essentiel pour plusieurs espèces bactériennes. À titre d'exemple, la bactérie intracellulaire facultative *Legionella pneumophila* utilise le SST4 Dot/Icm pour sécréter près de 300 effecteurs potentiels afin de survivre et de se multiplier à l'intérieur des macrophages (133). Tout comme le système de sécrétion de type III, le SST4 permet la sécrétion d'effecteurs du cytoplasme bactérien au cytosol de la cellule eucaryote cible. Une seule exception à cette règle a été décrite, soit pour la toxine pertussis de type AB qui est sécrétée à partir du périplasme par le système Ptl retrouvé chez *Bordetella pertussis* (134). Le système commun aux bactéries Gram négatif comprend un minimum de douze composants structuraux et leur nom, VirB1 à 11 et VirD4, se base sur l'étude du SST4 de *Agrobacterium tumefaciens* (135). Au niveau de la membrane interne, on retrouve trois composants qui forment le corps du complexe, soit VirB6, 8 et 10. Ce complexe est aussi associé à trois ATPases requises pour fournir l'énergie, soit VirB4, 11 et VirD4 (136). Deux lipoprotéines, VirB7 et 9, sont impliquées dans la formation d'un pore au niveau de la membrane externe (137, 138). Toutefois, le pore est directement formé par

le domaine C-terminal de VirB10 (139). Finalement, VirB2 et VirB5 constituent les pilines majeures et mineures, respectivement, du pilus extracellulaire associé au SST4 (140, 141). Deux récentes structures du complexe VirB7-9-10 (TraN, O et F) de *E. coli* ont été publiées, fournissant ainsi une meilleure image du complexe formé par le SST4 (139, 142). Ce dernier forme plusieurs anneaux distincts, chacun composé de 14 sous-unités protéiques, traversant ainsi l'enveloppe cellulaire, de la membrane interne à la membrane externe. Ce complexe formerait le canal par lequel passe les substrats lors du processus de sécrétion. Le rôle exact du pilus du SST4 est toujours sujet de débat. Une fonction de canal permettant le passage des substrats ainsi qu'une fonction dans l'adhésion lui ont déjà été attribuées (143, 144). Toutefois, certaines mutations dans les composants du système permettent de séparer la présence du pilus et la sécrétion d'effecteurs, suggérant ainsi que le pilus n'est pas essentiel pour la sécrétion (145).

#### 1.4.5. Système de sécrétion de type V

Le système de sécrétion de type V (SST5) est, selon moi, le plus simple de tous les systèmes de sécrétion présents chez les bactéries. Avec plusieurs centaines de membres identifiés, les protéines sécrétées via le système de sécrétion de type V représentent une des plus importantes familles de facteurs de virulence bactériens (Figure 4). Trois sous-familles font historiquement partie de ce système, soit les autotransporteurs monomériques (type Va), les autotransporteurs trimériques (type Vc) et les systèmes de sécrétion à deux partenaires (TPS, type Vb) (3). Toutefois, les auteurs d'une revue publiée récemment ont refait la classification des systèmes de sécrétion de type V pour y inclure deux nouvelles sous-familles, les type Vd et Ve. Une brève description de ces sous-familles sera faite dans cette section, excluant la famille des autotransporteurs monomériques puisqu'une section complète lui est dédiée (voir section 3 - autotransporteurs monomériques).



**Figure 4 : Système de sécrétion de type V** [adaptée de la figure 1a de la référence (146)]. Représentation schématique des différentes sous-familles du système de sécrétion de type V. Brun, domaine  $\beta$  ou domaine de translocation; orange, domaine périplasmique; vert pâle, domaine de jonction ou domaine TPS; vert foncé, domaine passager.

Les autotransporteurs trimériques (type Vc) sont retrouvés chez des espèces bactériennes d'importance médicale, telles que *E. coli*, *Haemophilus*, *Neisseria meningitidis*, *Yersinia pestis*, parmi plusieurs autres (147), ainsi que chez des espèces commensales ou retrouvées dans l'environnement. Tous les autotransporteurs trimériques décrits sont des protéines localisées dans la membrane externe avec comme fonction prédominante, l'adhésion (148). Une variété de ligands pour cette famille de protéines a été identifiée, incluant des récepteurs cellulaires spécifiques ainsi que des protéines de la matrice extracellulaire. De plus, certaines de ces protéines sont aussi importantes pour moduler la réponse immunitaire de l'hôte (149) ou sont impliquées dans des phénotypes tels que la formation de biofilm et l'autoagglutination (150).

Les autotransporteurs trimériques sont des protéines homotrimériques obligatoires, ce qui signifie que l'oligomérisation est requise pour leur localisation, stabilité et fonctionnalité (151). Leur structure générale correspond à un domaine C-terminal extrêmement bien conservé ainsi qu'un domaine passager N-terminal de longueur variable, exposé dans le milieu extracellulaire et responsable de la fonctionnalité de la protéine. Le domaine C-terminal correspond à un tonneau  $\beta$  formé de 12 brins  $\beta$ , où chaque monomère contribue pour 4 brins  $\beta$  (152). Ce tonneau va permettre le passage du domaine extracellulaire à travers la membrane externe (153). Le mécanisme exact de sécrétion est toujours un sujet de débat et sera abordé en détail dans la section des autotransporteurs monomériques. Toutefois, le modèle général suggéré est que les chaînes polypeptidiques correspondant au

domaine passager traversent la membrane externe via le pore formé par le tonneau  $\beta$ , et ce, avec l'aide du complexe Bam (146).

L'organisation structurale du domaine extracellulaire est conservée et correspond à une structure modulaire, les principaux modules étant la tige et la tête (154). La tige forme une structure étendue (coiled-coil) qui permet de positionner la tête loin de la membrane bactérienne (155). La structure de la tête de YadA a été la première résolue et correspond à un rouleau  $\beta$  gauche comportant 9 tours (156). Depuis, plusieurs autres structures ont été publiées montrant des variations dans le domaine de tête, malgré une forte prévalence de structures riches en brins  $\beta$  (147). Ce module de tête riche en brin  $\beta$  est retrouvé chez tous les autotransporteurs trimériques et peut même être présent en plusieurs copies, comme c'est le cas pour BpaA retrouvé chez *Burkholderia pseudomallei* (157, 158).

Quel est l'avantage d'exposer trois copies d'une même protéine, liées de façon covalente, à la surface de la bactérie? Cette caractéristique structurale propre aux autotransporteurs trimériques peut servir de plusieurs façons. L'architecture trimérique de ces protéines confère une résistance accrue aux protéases ainsi qu'à divers détergents, augmentant ainsi leur stabilité. De plus, cette architecture particulière permet à la bactérie d'exposer trois faces identiques de la protéine, permettant ainsi d'avoir plusieurs domaines adhésion similaires sur une même molécule. Cette caractéristique suggère que la protéine peut établir une interaction multivalente avec son récepteur et, par conséquent, augmenter l'affinité et la stabilité de cette interaction.

La principale caractéristique des systèmes de sécrétion à deux partenaires (TPS, Type Vb) est que la sécrétion de la protéine effectrice (TpsA) requière une seconde protéine (TpsB) pour le passage à travers la membrane externe (159). En fait, l'organisation des systèmes TPS est très similaire à celle des autotransporteurs, à l'exception que le domaine passager et le domaine  $\beta$  se retrouvent sur deux polypeptides différents. Les deux principaux systèmes TPS étudiés sont FHA/FhaC, une hémagglutinine filamenteuse de *Bordetella pertussis*, et l'adhésine HMW1/HMW1B de *Haemophilus influenzae*. Malgré le fait que ces deux protéines modèles font partie de la même famille, leur biogenèse comporte quelques différences majeures qui méritent d'être discutées.

De façon générale, les protéines TpsA ont des caractéristiques structurales assez bien conservées; ce sont des protéines de grande taille (souvent plus de 3 000 acides aminés), leur séquence est formée de motifs répétés et elles adoptent préférentiellement une structure



en forme d'hélice  $\beta$  (157). Plusieurs fonctions ont été décrites pour les protéines TpsA, telles qu'une implication dans l'adhésion et l'invasion de cellules eucaryotes, une action cytotoxique ou un rôle direct dans l'inhibition de la croissance de façon contact dépendant (159, 160). Toutes les protéines TpsA possèdent une séquence signal permettant leur transport dans le périplasma via le système Sec. Le domaine TPS ou domaine de sécrétion est une région N-terminale d'environ 300 acides aminés qui est retrouvée chez toutes les protéines TpsA (161, 162). Ce domaine TPS est nécessaire et suffisant pour permettre l'interaction avec le TpsB et promouvoir la sécrétion à travers la membrane externe.

Au delà de ces similarités, beaucoup de différences persistent entre les deux TpsA prototypes, soit FHA et HMW1 (163). FHA possède un domaine TPS en N-terminal et un pro-domaine, d'environ 1200 acides aminés localisé à l'extrémité C-terminale, qui est clivé lors du processus de sécrétion (164). Le rôle de ce dernier demeure toujours inconnu, quoique plusieurs hypothèses existent, telles que la prévention contre une dégradation prématurée de FHA mature dans le périplasma ou pour moduler la conformation finale de FHA mature (165). Côté topologie, le modèle mis de l'avant suggère que la partie C-terminale de FHA mature est exposée à la surface de la bactérie, ce qui concorde avec la fonction d'adhésion de ce domaine (165). HMW1 est aussi synthétisée sous la forme d'une pré-pro-protéine (166). Toutefois, le pro-domaine d'environ 360 acides aminés est localisé à l'extrémité N-terminale et inclus le domaine TPS. Contrairement à FHA, la topologie de HMW1 suggérée prédit que l'extrémité N-terminale, après le clivage du pro-domaine, est exposée à la surface. L'extrémité C-terminale reste du côté périplasmique et forme une boucle grâce à la présence de deux résidus cystéines. Cette boucle empêche la sécrétion, bloque le pore formé par HMW1B et permet l'ancrage de l'adhésine HMW1 à la surface de la bactérie (167).

Les protéines TpsB font partie de la superfamille TpsB/BamA des protéines de membrane externe (168). La structure de FhaC, le TpsB permettant la sécrétion de FHA, a récemment été résolue (169). FhaC forme un tonneau  $\beta$  monomérique composé de 16 brins  $\beta$  antiparallèles et possède un module périplasmique essentiel à la sécrétion, contenant deux domaines POTRA. À l'opposé, des études de microscopie électronique montrent que la protéine TpsB HMW1 se retrouve dans la membrane externe sous forme de dimère (170). De plus, HMW1B formerait un tonneau  $\beta$  composé de seulement 10 brins  $\beta$  et posséderait un domaine périplasmique contenant un domaine POTRA unique (171). Toutefois, le point commun de ces systèmes TPS est qu'ils semblent utiliser une voie de sécrétion plus

« indépendante » que celle requise pour la sécrétion des autotransporteurs. En effet, une étude récente faite avec FHA a démontré que le TpsB seul permettait la sécrétion *in vitro* de FHA, sans avoir recours à des facteurs accessoires comme le complexe Bam (172). Les protéines TpsB pourraient donc servir de modèle simplifié pour l'étude de la famille de protéines TpsB/BamA. Il est toutefois important de noter que dans un contexte *in vivo*, la sécrétion des protéines TpsA requière aussi la présence de chaperonnes périplasmiques telles que DegP (173).

L'identification de membres appartenant au système de sécrétion de type V est basée sur trois aspects communs : la présence d'une séquence signal reconnue par le système Sec, d'un tonneau  $\beta$  nécessaire pour l'insertion dans la membrane externe ainsi que la présence d'un domaine passager fonctionnel exposé à la surface de la bactérie ou sécrété dans le milieu extracellulaire. Récemment, les auteurs d'une revue sur le système de sécrétion de type V ont suggéré, basé sur les critères mentionnés ci-dessus, l'introduction de deux nouvelles sous-familles (Vd et Ve) (146). Il faut toutefois mentionner que cette nouvelle nomenclature n'est pas encore adoptée dans la littérature vue la récente date de publication. Le type Vd est représenté par la protéine PlpD (patatin-like protein), retrouvée chez *Pseudomonas aeruginosa* (174). Cette protéine a une organisation unique puisqu'elle semble correspondre à une fusion entre un autotransporteur monomérique et un système TPS. Plus précisément, PlpD possède une séquence signal permettant son transport par le système Sec suivi par un domaine passager qui possède une fonction enzymatique et un domaine  $\beta$  à l'extrémité C-terminale. La nouveauté est que le domaine  $\beta$  est formé de 16 brins  $\beta$ , typique des systèmes TPS, ainsi que d'un domaine POTRA (174). La présence d'un domaine POTRA est unique et ouvre la porte à de nombreuses questions. Si les systèmes TPS sont indépendants et ne requièrent pas l'aide du complexe Bam pour le passage à travers la membrane externe, qu'en est-il de cet autotransporteur possédant un domaine POTRA? Est-ce qu'il est sécrété indépendamment de BamA ou le domaine POTRA interne est requis pour le passage du domaine passager ? Il existe plus de 200 orthologues de PlpD, retrouvés autant dans des souches d'importance médicales que environnementales, ce qui permet de délimiter une nouvelle sous-famille de système de sécrétion de type V. Toutefois, une caractérisation plus approfondie de PlpD est requise pour pouvoir commencer à discuter du mode de sécrétion de cette sous-famille.

La deuxième nouvelle sous-famille du Type V, le Type Ve, représente des autotransporteurs avec une organisation inversée (146). En fait, les protéines classées dans le type Ve ne correspondent pas à des protéines nouvellement découvertes, mais plutôt des protéines pour lesquelles de nouvelles informations permettent maintenant de les relier à la famille des autotransporteurs. Les principaux membres sont ceux appartenant à la famille intimin/invasin (175). Cette famille de protéines possède plus de 69 homologues dans différentes espèces de  $\gamma$ ,  $\beta$  et  $\epsilon$ -proteobacteria ainsi que chez certaines espèces de *Chlamydia* (175). Le modèle topologique suggéré pour ces protéines correspond à un domaine périplasmique (possiblement connecté au peptidoglycan), un tonneau  $\beta$  possédant entre 10 et 16 segments transmembranaires prédits, une hélice  $\alpha$  localisée à l'intérieur du pore formé par le tonneau  $\beta$  et un domaine extracellulaire portant la fonction. Ce modèle est très similaire à celui des autotransporteurs monomériques, à l'exception du domaine périplasmique. La principale différence est l'organisation du polypeptide. Contrairement aux autotransporteurs classiques, la région N-terminale correspond au domaine  $\beta$  et la région C-terminale au domaine passager, suggérant un mode de sécrétion légèrement différent et une topologie finale inversée. Il faut toutefois mentionner que les auteurs de l'étude sur la topologie des protéines de type intimine/invasin ont décrit ces protéines comme une nouvelle famille de protéines sécrétées plutôt que comme une nouvelle classe d'autotransporteurs (175). Il reste maintenant à voir si cette nouvelle nomenclature va être acceptée et utilisée à l'avenir dans la littérature.

#### 1.4.6. Système de sécrétion de type VI

Le système de sécrétion de type VI (SST6) est le moins bien connu des systèmes décrits ici, principalement du fait qu'il a été découvert assez récemment (176). Ce système se caractérise par une organisation génétique assez bien conservée correspondant à un groupe de 15 à 20 gènes, dont 13 seraient essentiels pour un système fonctionnel. On retrouve ce groupe de gènes chez environ 25% des bactéries Gram négatif, majoritairement chez les protéobactéries (177). Plusieurs rôles ont été attribués à ce système, notamment l'induction d'une toxicité au niveau des amibes et cellules phagocytaires aussi bien qu'auprès d'espèces bactériennes différentes, en plus d'une implication dans la formation des biofilms (176, 178, 179).

Plusieurs composants structuraux ont été identifiés au cours des dernières années. Deux composants structuraux majeurs sont essentiels à la sécrétion, soit la protéine Hcp (Haemolysin co-regulated protein) et les protéines VgrG (Valine-Glycine repeat protein G). Hcp forme des anneaux hexamériques qui ont la capacité de polymériser *in vitro* (180). Les protéines VgrG présentent une forte similarité avec la queue du bactériophage T4, requise pour perforer la membrane bactérienne (181). Le mode d'action généralement suggéré pour ces protéines consiste en la formation d'un trimère de VgrG formant ainsi une structure similaire à une seringue. Cette seringue se situerait au sommet d'une structure similaire à un pilus, formée par l'oligomérisation de Hcp. Bien que ce modèle soit attrayant, aucune interaction directe n'a encore été démontrée entre ces deux protéines.

Les protéines de membrane interne IcmF, DotU (IcmH) ainsi qu'une lipoprotéine associée à la membrane externe, SciN sont aussi des composants requis pour le fonctionnement du système (182, 183). Un fait intéressant à noter est que les protéines IcmF et DotU sont aussi des protéines accessoires du SST4, ce qui signifie que le SST6 utilise des unités fonctionnelles d'autres systèmes de sécrétion, une particularité plutôt unique (184). Deux autres protéines récemment découvertes, soit VipA et VipB, permettent d'émettre de nouvelles hypothèses sur le mode d'action du SST6 (185). Ces deux protéines forment des tubes possédant un diamètre interne de 100 Å, assez large pour contenir la structure de type pilus générée par la protéine Hcp (diamètre externe de 85 Å). Une particularité de VipA/B est que l'hydrolyse de l'ATP par une ATPase essentielle au SST6, ClpV, entraîne la dissociation des ces tubes. Le modèle proposé par une étude récente est donc que le SST6 serait un système d'injection contractile, où la contraction de VipA/B fournirait l'énergie et la force requise pour propulser le complexe Hcp/VgrG à travers la membrane de la cellule cible (186). ClpV permettrait ensuite la dépolymérisation du complexe VipA/B, permettant ainsi leur recyclage.

Quels sont les effecteurs sécrétés par le SST6? Cette question est primordiale puisqu'on ignore toujours comment ce système permet la sécrétion de protéines solubles. Un premier groupe d'effecteurs est représenté par quelques unes des protéines VgrG. Ces dernières possèdent une extension C-terminale avec un domaine effecteur, suffisant pour causer une cytotoxicité (181). Dans cette optique, la seringue du système servirait à perforer la membrane cible en plus d'agir à titre effecteur. Toutefois, d'autres effecteurs bactériens sans séquence signal Sec-dépendante, comme la toxine Tse2 de *Pseudomonas aeruginosa*, sont sécrétés de façon dépendante du SST6 (178). Dans ce cas, les protéines sécrétées

doivent traverser le conduit formé par Hcp et VgrG. Le diamètre interne de 40Å du canal formé de Hcp serait suffisant pour permettre le passage du polypeptide déplié. Par contre, le conduit formé par VgrG (environ 12Å, basé sur le bactériophage T4) serait probablement trop étroit. Une des hypothèses envisagées est que le complexe VgrG se dissocie du pilus Hcp suite à la perforation de la membrane cible de façon à permettre la translocation des effecteurs (187). Encore une fois, ce modèle repose sur très peu d'évidence et beaucoup reste à faire pour comprendre ce nouveau complexe de sécrétion qui traverse l'enveloppe bactérienne.

Ce bref résumé sur les systèmes de sécrétion permet de mieux reconnaître la complexité des systèmes présents chez les bactéries et de comprendre l'importance de leur étude. Si les bactéries ont acquis autant de moyens pour permettre aux protéines d'être exposées à leur surface ou d'être sécrétées, c'est que la fonction jouée par ces molécules effectrices est primordiale pour la survie des bactéries, autant dans l'environnement qu'au niveau d'un organisme hôte.

## **2. La glycosylation chez les bactéries**

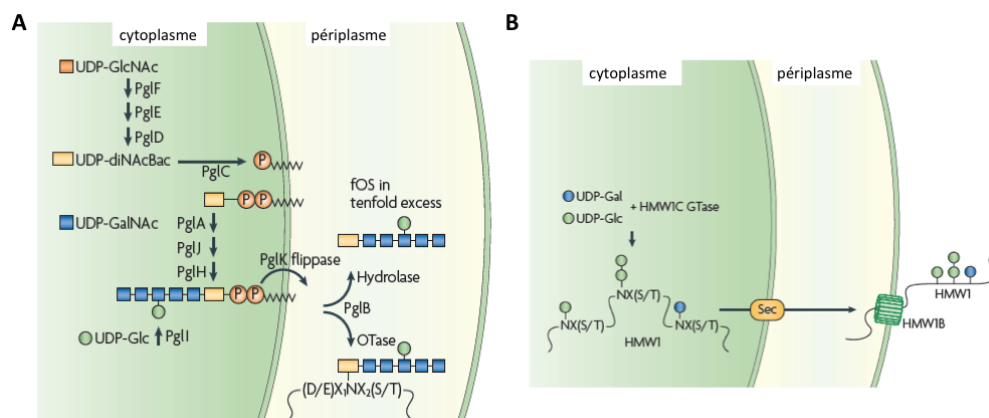
Les sucres sont des constituants essentiels des bactéries. On les retrouve particulièrement au niveau des structures de la membrane bactérienne telles que les lipopolysaccharides (LPS) ou les capsules (188, 189). Leur synthèse et assemblage ainsi que leur rôle sont maintenant bien connus. Ces composants sont essentiels pour plusieurs processus cellulaires, dont le maintien de l'intégrité membranaire, la protection contre l'environnement, la modulation de la réponse de l'hôte, parmi d'autres. Au cours des dernières années, il est toutefois devenu apparent qu'au delà de ces structures classiques, une variété de protéines bactériennes sont aussi modifiées par l'ajout de sucres et ce, pour une grande diversité d'espèces bactériennes. À titre d'exemple, plus de 1000 protéines potentiellement glycosylées ont récemment été identifiées chez *Bacteroides fragilis*, incluant des protéines sécrétées ainsi que des protéines impliquées dans des processus fondamentaux tels que la division cellulaire (190).

Pourquoi modifier une protéine par l'ajout de sucres? La glycosylation est une modification très courante chez les eucaryotes et a été impliquée dans une multitude de processus cellulaires. Chez les bactéries, plusieurs fonctions ont aussi été identifiées (191). L'ajout de

sucres sur une protéine peut modifier les propriétés physico-chimiques, telles que la solubilité ou la charge globale. La glycosylation a aussi été impliquée dans le maintien d'une conformation adéquate, dans l'augmentation de la stabilité ainsi que pour la protection de la protéine contre une dégradation prématurée. Finalement, les glycoprotéines exposées à la surface de la bactérie peuvent aussi être directement impliquées dans l'interaction avec les cellules hôtes ou dans l'interaction bactérie-bactérie.

## 2.1. N-glycosylation

Deux systèmes majeurs de N-glycosylation sont étudiés chez les bactéries. Le premier correspond au système général de N-glycosylation mis en évidence chez *Campylobacter jejuni* (Figure 5A). Le second système, beaucoup plus récent et bien représenté par la protéine HMW1C de *Haemophilus influenzae*, implique seulement une glycosyltransférase dite atypique (Figure 5B). Ces deux systèmes seront décrits dans les prochaines sections. La N-glycosylation chez les bactéries Gram positif ne sera pas discutée puisqu'il a peu d'exemples disponibles dans la littérature et aucun système général ou répandu n'a été décrit à ce jour (191).



**Figure 5 : N-glycosylation chez les bactéries Gram négatif** [figure adaptée de la référence (192)]. **A.** Système général de N-glycosylation chez *C. jejuni*. Les sucres sont synthétisés dans le cytoplasme en association avec un précurseur lipidique, le Und-PP. L'heptasaccharide complet est transporté dans le périplasme par la flippase spécifique PglK, puis transféré sur le groupement amine d'une asparagine par l'oligosaccharyltransférase PglB ou relâché dans le périplasme sous la forme de fos (oligosaccharides libres). **B.** N-glycosylation de HMW1 chez *H. influenzae*. L'oligosaccharyltransférase HMW1c transfère un sucre (glucose ou galactose) sur site consensus de HMW1 dans le cytoplasme avant la sécrétion par le système Sec. HMW1c permet aussi la formation de lien hexose-hexose.

### 2.1.1. Système général de N-glycosylation

Le premier système général de N-glycosylation a été mis en évidence chez la protéobactérie epsilon *Campylobacter jejuni* (193). Ce système reste toujours le mieux caractérisé et sert de modèle d'étude, malgré le fait que plusieurs autres protéobactéries epsilon et protéobactéries delta possèdent des homologues de cette voie de glycosylation (192). On considère que ce système est général puisque près de 150 protéines de *C. jejuni* sont prédites pour être glycosylées, une confirmation expérimentale ayant été faite sur plus de 65 de ces protéines (194). Le terme N-glycosylation signifie que le sucre est attaché de façon covalente à un résidu asparagine. La séquence consensus pour la N-glycosylation bactérienne a été identifiée et correspond à Asp/Glu-X-Asn-X-Ser/Thr, où X représente n'importe quel acide aminé à l'exception de la proline. Ce motif est un peu plus complexe que celui reconnu par la machinerie de glycosylation eucaryote (Asn-X-Ser/Thr) (195). Il faut toutefois noter que certaines variations de ce motif ont aussi été décrites récemment (196).

Les gènes requis pour permettre la glycosylation se situent dans un locus de 16 kb nommé *pgl* (protein glycosylation). La Figure 5A montre la voie de glycosylation générale utilisée par *C. jejuni* (192). Globalement, le sucre utilisé pour modifier les protéines cibles, soit un heptasaccharide (GalNAc<sub>2</sub>[Glc]GalNAc<sub>3</sub>-diNAcBac), est synthétisé dans le cytoplasme de la bactérie, en association avec un précurseur lipidique, le undecaprenyl pyrophosphate (Und-PP). Le sucre est ensuite transféré dans le périplasme par l'action d'une flippase ATP-dépendante spécifique, soit PglK. C'est l'oligosaccharyltransférase PglB qui va transférer le sucre du précurseur lipidique à la protéine cible au niveau de la séquence consensus (197). PglB permet le transfert du sucre de façon post-traductionnelle sur des protéines repliées au niveau de régions exposées et structurellement flexibles (198). Le système de N-glycosylation chez les eucaryotes est assez similaire, ce qui concorde avec l'homologie existante entre l'enzyme PglB et la sous-unité STT3 du complexe oligosaccharyltransférase eucaryote (199). Les principes directeurs tels que la synthèse des sucres sur un précurseur lipidique et le transfert des sucres sur les protéines du côté interne du réticulum endoplasmique (même sens topologique que la glycosylation périplasmique chez les bactéries) sont bien conservés entre les différents domaines de la vie. Toutefois, chez les eucaryotes, la N-glycosylation est couplée au processus de sécrétion et se fait de manière co-traductionnelle sur des polypeptides majoritairement dépliés.

L'enzyme PglB est aussi une protéine bi-fonctionnelle qui possède une activité d'hydrolase (200). Cela signifie que PglB peut, en plus de transférer des sucres sur une protéine, entraîner la relâche de sucres libres dans le milieu périplasmique (soit pour free oligosaccharides). En fait, la quantité d'heptasaccharides générée par la voie *pgl* serait 10 fois supérieure sous forme libre périplasmique que sous la forme attachée aux protéines (201). D'ailleurs, toutes les protéobactéries étudiées possèdent une importante quantité de sucres périplasmiques variables nommés OPG (osmoregulated periplasmic glucans). Ces sucres peuvent atteindre jusqu'à 5 à 7 % du poids sec total chez *E. coli*, selon les conditions de croissance (202). La principale fonction attribuée à ces OPG est de réguler l'effet d'une variation dans l'osmolarité en influençant le volume et la densité du périplasma (203). Une mutation dans les voies de biosynthèse responsables de la production des OPG entraîne des phénotypes multiples, incluant une hyperproduction des exopolysaccharides de surface, un défaut de motilité ainsi qu'une hypersensibilité aux antibiotiques. L'effet d'une inactivation du système *pgl* chez *C. jejuni* cause une diminution importante de la virulence autant *in vitro* que *in vivo* dans un modèle de colonisation du poulet (204). Toutefois, considérant le fait que la présence des OPG est essentielle pour la virulence de plusieurs pathogènes bactériens (203), il est maintenant primordial de déterminer si l'effet de l'abolition de la voie *pgl* est dû à une absence de glycosylation protéique et/ou à une absence des fos dans le périplasma.

La N-glycosylation a initialement été découverte chez les archaebactéries (205). Chez ces organismes, la glycosylation est beaucoup plus fréquente, modifiant une variété de protéines dont la flagelline et les protéines de l'enveloppe (S-layers). Peu d'études ont vérifié le rôle exact de ces modifications, mais l'hypothèse mise de l'avant est que la glycosylation aurait une fonction possible dans la capacité des archaebactéries à vivre dans des conditions de croissance extrême (206).

### 2.1.2. N-glycosyltransférases atypiques

Le deuxième système de N-glycosylation bactérien correspond à la famille de glycosyltransférases représentée par HMW1C de *H. influenzae* (207, 208) (Figure 5B). Cette oligosaccharyltransférase est dite atypique puisqu'elle ne suit pas les règles communes à la N-glycosylation chez les autres organismes vivants (192). Tout d'abord, l'ajout des sucres sur la protéine cible ne requiert qu'une seule protéine, soit HMW1C ou

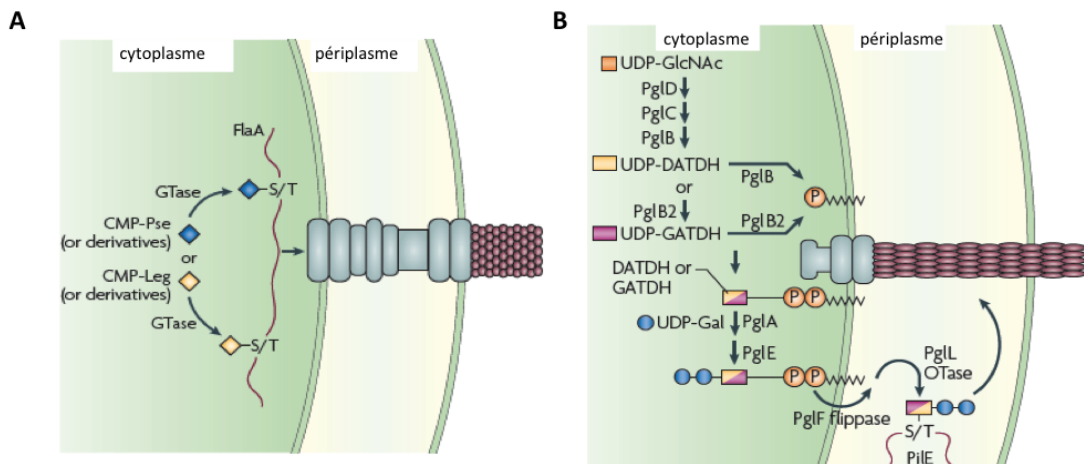


un de ses homologues, contrairement au complexe multi-enzymatique retrouvé chez *C. jejuni*. Les sucres ajoutés ne sont pas des sucres complexes spécifiques, mais plutôt des résidus glucose ou galactose détournés de la voie de biosynthèse des lipooligosaccharides (LOS) (209). HMW1C reconnaît le motif consensus simplifié de N-glycosylation chez les eucaryotes, soit Asn-X-Ser/Thr, et ne semble pas nécessiter la présence d'un résidu chargé négativement en position -2 (207). HMW1C, en plus d'agir en tant N-glycosyltransférase, peut aussi faire des liaisons hexose-hexose directement sur la protéine cible, lui conférant ainsi une double activité (207). Finalement, HMW1C est une protéine cytoplasmique, ce qui signifie que la glycosylation se fait avant le transport à travers la membrane interne, une caractéristique unique chez les bactéries Gram négatif (209). L'ensemble de ces caractéristiques, qui sont bien différentes de celles observées pour le système général de N-glycosylation, a permis de définir un nouveau type de N-glycosyltransférases bactériennes. Toutefois, cette famille de glycosyltransférases est plutôt restreinte, incluant à ce jour uniquement 8 membres potentiels (207). HMW1C et ses homologues se différencient aussi du système général de N-glycosylation puisqu'ils ne ciblent qu'une seule protéine. En effet, HMW1C de *H. influenzae* fait partie d'un opéron de trois gènes codant aussi pour les protéines HMW1A et HMW1B, un système de sécrétion à deux partenaires (TPS) (209). Cette association avec TPS est une particularité conservée chez tous les homologues de HMW1C. La glycosylation de HMW1 est requise pour prévenir une dégradation prématurée ainsi que pour permettre l'ancrage de cette adhésine à la surface bactérienne (209). La plupart des protéines sécrétées par un TPS ont des tailles anormalement grandes, comme c'est le cas de HMW1. Par conséquent, cette N-glycosylation atypique pourrait être une caractéristique spécifique à certaines protéines sécrétées par le système de sécrétion de type V afin d'optimiser l'expression et la stabilité de ces dernières à la surface de la bactérie.

## 2.2. O-glycosylation

Plusieurs systèmes de O-glycosylation chez les bactéries ont été décrits jusqu'à maintenant, incluant des systèmes généraux comme ceux retrouvés chez *Neisseria meningitidis* ou *Bacteroides fragilis*, mais aussi des systèmes plus spécifiques dont ceux reliés à la glycosylation des pili de type IV, des flagelles, des autotransporteurs et des adhésines

riches en résidus sérine (Figure 6). Ces systèmes de O-glycosylation bactériens vont être décrits dans les prochaines sections.



**Figure 6 : O-glycosylation chez les bactéries Gram négatif** [adaptée de la référence (192)]. **A.** O-glycosylation des flagelles chez *C. jejuni*. Les protéines structurales formant le filament sont possiblement glycosylées à l'interface entre la membrane interne et le cytoplasme. Les sucres (des dérivés de l'acide pseudaminic ou legionaminic) sont ajoutés individuellement sur des résidus sérine. **B.** O-glycosylation des pili de type IV chez *N. meningitidis*. Les sucres sont assemblés au niveau de la face cytoplasmique de la membrane interne puis transférés dans le périplasme par la flippase PglF. Dans le périplasme, l'oligosaccharyltransférase ajoute les sucres sur un résidu sérine ou thréonine de la piline (PilE) ou d'une autre protéine périplasmique ou membranaire.

### 2.2.1. O-glycosylation des flagelles

La glycosylation de la sous-unité structurale majeure du filament qui compose le flagelle est maintenant bien documentée pour plusieurs espèces bactériennes, telles que *C. jejuni*, *Helicobacter pylori*, *Pseudomonas aeruginosa*, *Clostridium* et *Listeria monocytogenes* (210). Le terme O-glycosylation réfère à l'ajout d'un sucre sur un résidu sérine ou thréonine, avec quelques exceptions pour lesquelles la tyrosine peut aussi être modifiée. Comme pour le système général de N-glycosylation, *C. jejuni* est un modèle d'étude pour la O-glycosylation de la flagelline. Chez cette bactérie, FlaA et FlaB sont les deux protéines structurales qui forment le filament (211). Les études portant sur FlaA ont démontré que 19 sites (sérine ou thréonine) sont modifiés par des sucres variables, majoritairement des dérivés de l'acide pseudaminic ou legionaminic. La quantité de sucres retrouvée sur la flagelline peut atteindre jusqu'à 10% de la masse totale de cette dernière, illustrant bien l'importance de cette modification. Côté fonctionnalité, la glycosylation serait requise pour

l'assemblage adéquat du filament flagellaire, et donc requis pour la motilité (212). Toutefois, seulement certains sites de glycosylation sur FlaA sont réellement essentiels pour l'assemblage. En effet, pour cinq de ces 19 sites identifiés, l'absence de glycosylation ne nuit pas à la stabilité, mais plutôt à la capacité du filament à auto-agglutiner, suggérant ainsi un double rôle pour la glycosylation (213).

Un locus de plus de 50 gènes a été identifié comme responsable de la synthèse des sucres destinés à FlaA et FlaB, expliquant ainsi la variabilité observée quant à la structure exacte du sucre (214). Malgré la connaissance de ce locus génétique, le mécanisme de glycosylation reste indéterminé et le modèle actuellement mis de l'avant se base plus sur des déductions que sur des preuves expérimentales (Fig. 6A). Tout d'abord, les sous-unités du filament doivent être sécrétées à la surface par le SST3 dédié à la synthèse du flagelle. Les protéines substrats sont amenées au corps basal du système de sécrétion via une chaperonne spécifique puis sécrétées jusqu'à la surface, à l'extrémité croissance du filament, sans contact avec le milieu périplasmique (215). Il est assez difficile de croire que la glycosylation puisse avoir lieu au niveau de la surface bactérienne et les flagellines n'ont pas accès à l'espace périplasmique, suggérant une glycosylation cytoplasmique. Toutefois, pour la plupart de ces systèmes de O-glycosylation, les oligosaccharyltransférases responsables du transfert du sucre à la flagelline ne sont pas connues, ce qui signifie que leur localisation précise ne peut pas être déterminée. Le modèle le plus plausible suggère donc que les sous-unités du filament vont être modifiées dans l'environnement restreint du corps basal par une glycosyltransférase associée à la face cytoplasmique de la membrane interne ou à un composant du SST3.

Ce modèle est attrayant mais génère aussi beaucoup de questions sur le mécanisme de reconnaissance précis des sites de glycosylation. Aucune séquence consensus, qui expliquerait pourquoi ce sont toujours les mêmes sites qui sont modifiés sur FlaA, n'a pu être déterminée. Basé sur une modélisation de la structure du filament de *C. jejuni*, il a été suggéré que les sucres se retrouvent sur les surfaces exposées du filament (216). Toutefois, cela signifie que la glycosyltransférase doit reconnaître un certain niveau de structure dans le polypeptide cytoplasmique pour justifier une localisation précise dans la protéine complètement repliée. Une autre perspective relative à ce modèle est que la présence d'une glycosyltransférase associée à un SST3 entraîne la possibilité que d'autres molécules effectrices utilisant cette voie de sécrétion puissent aussi être modifiées, quoique au meilleur de mes connaissances, je ne crois pas qu'un effecteur du SST3 ait déjà été décrit

comme glycosylé. Donc, il est assez évident que beaucoup de travail reste à faire pour bien comprendre le mécanisme de glycosylation spécifique aux sous-unités composant le filament des flagelles.

### 2.2.2. O-glycosylation des pili de type IV

Le pili de type IV est retrouvé chez plusieurs espèces de bactéries pathogènes dont *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Francisella tularensis* et plusieurs souches de *E. coli* entéropathogénique (EPEC). Ces pili sont impliqués directement dans la virulence et sont importants pour la motilité, la susceptibilité aux bactériophages, la colonisation des muqueuses ainsi que pour la conjugaison et l'acquisition d'ADN dans le cas des bactéries naturellement transformables (217). Pour quelques unes de ces espèces bactériennes, un système de glycosylation spécifique aux pili de type IV a été identifié. Les plus connus sont ceux de *N. meningitidis* et *P. aeruginosa*.

Le mécanisme de O-glycosylation des pilines est très similaire au mécanisme de N-glycosylation général retrouvé chez *C. jejuni* (Figure 6B) (218). Un complexe d'enzymes cytoplasmiques est responsable de la synthèse du sucre, qui en association avec un Und-PP va être transféré du côté périplasmique de la membrane interne par une flippase spécifique. Dans le périplasme, une oligosaccharyltransférase va ajouter les sucres sur les sous-unités du pili de type IV. Dans le cas de *N. meningitidis*, certaines enzymes impliquées dans la synthèse du sucre sont soumises à une variation de phase, ce qui influence la longueur et la nature exacte du sucre ajouté (219).

Des différences entre et au sein des espèces existent pour ce qui est de la localisation des sucres sur les pilines. Chez *P. aeruginosa* 1244, un seul sucre est lié à la sérine C-terminale, tandis que pour la souche PA5196, les pilines sont modifiées sur plusieurs sites par une glycosyltransférase unique à cette souche (220, 221). Pour *N. gonorrhoeae*, la sérine en position 63 est le seul résidu accepteur identifié (222). Un point intéressant à noter est que le pilus retrouvé chez *N. gonorrhoeae* peut aussi être modifié par l'addition de phosphoéthalonamine ou de phosphocholine et ce, sur les sites possiblement occupés par les sucres (223, 224). Cette nouvelle modification, en plus de la variation de phase qui amène une diversité dans la structure même du sucre, pourrait être un moyen employé par la bactérie pour modifier les propriétés antigéniques de ses structures extracellulaires.

Le rôle de la glycosylation des pili est toujours incertain. Chez *P. aeruginosa* 1244, une mutation de l'oligosaccharyltransférase PilO cause une petite diminution de la motilité ainsi qu'un désavantage pour la colonisation des voies respiratoires dans un modèle d'infection chez la souris, lorsqu'il y a compétition avec une souche sauvage (225). Toutefois, la morphologie des pili n'est pas affectée et la bactérie possède toujours la capacité de former des biofilms. Pour *N. meningitidis*, l'abolition de la glycosylation cause une légère augmentation du nombre de pili à la surface de la bactérie et, par conséquent, une augmentation de l'adhésion (226). Toutefois, une étude récente a démontré que le pili non-glycosylé confère un phénotype hyper-adhérent à *N. meningitidis* mais qu'il est cependant incapable de permettre l'invasion des cellules épithéliales cervicales (227).

### 2.2.3. O-glycosylation des autotransporteurs monomériques de *E. coli*

Il n'y a pas clairement de famille de O-glycosyltransférases chez *E. coli*. Toutefois, quatre autotransporteurs monomériques de *E. coli*, soit l'adhésine AIDA-I (Adhesin Involved in Diffuse Adherence), l'invasine TibA (enterotoxigenic invasion locus B), le facteur agrégatif Ag43 et la protéine EhaJ ont été décrits comme glycosylés (7-9, 228). AIDA-I, TibA et EhaJ sont retrouvés sous la forme d'un opéron comprenant deux gènes et suivant l'organisation suivante : *aah-aidA*, *tibC-tibA* et *ehaJ-egtA*. Aah, TibC et EgtA sont les glycosyltransférases spécifiques de AIDA-I, TibA et EhaJ, respectivement, et présentent une forte similarité de séquence. Dans le cas de l'Ag43, aucune glycosyltransférase spécifique n'a été identifiée, quoique cette protéine puisse être retrouvée sous forme glycosylée dans une souche sauvage de *E. coli* uropathogène et une souche isolée de matière fécale (9, 229). Il est à noter que la fonction des glycosyltransférases Aah et TibC est interchangeable, suggérant un mécanisme d'action similaire (230).

Le rôle de la glycosylation pour ces quatre autotransporteurs monomériques n'est pas clairement défini et varie selon la protéine étudiée. Pour AIDA-I et TibA, il a été démontré que la glycosylation est requise pour le phénotype d'adhésion (7, 8), mais dispensable pour l'agrégation entre bactéries et la formation de biofilm (231, 232). Pour EhaJ, la glycosylation permet une meilleure capacité à former des biofilms, mais n'est pas requise pour l'adhésion aux composants de la matrice extracellulaire (228). Dans le cas de l'Ag43, plusieurs études entrent en contradiction. Un point important à noter à ce sujet est que toutes les études portant sur la glycosylation de l'Ag43 ont été faites dans un contexte de

co-expression avec Aah. Étant donné que l'Ag43 n'est pas le substrat natif de Aah, la glycosylation peut ne pas être constante ou optimale. L'effet de la glycosylation dépend aussi beaucoup de l'allèle de l'Ag43 étudiée, ce qui complexifie l'interprétation (229). De façon générale, la glycosylation semble n'avoir aucun effet ou entraîner une diminution de la capacité de l'Ag43 à permettre l'agrégation entre bactéries (229, 233). De façon similaire, la formation de biofilm n'est généralement pas influencée par la présence de Aah. Pour le phénotype d'adhésion, lorsque l'allèle étudié permet l'adhésion aux cellules eucaryotes, la glycosylation semble avoir un effet négatif. À l'opposé, la présence des sucres optimiserait la capacité de l'Ag43 à interagir avec les composants de la matrice extracellulaire (229). Donc, malgré la forte similarité entre ces trois glycosyltransférases, l'impact de cette glycosylation sur les protéines cibles reste à éclaircir et semble varier entre les protéines et même entre différents allèles.

Le mode de fonctionnement des ces glycosyltransférases demeure inconnu. La protéine Aah (Autotransporter Adhesin Heptosyltransférase) utilise le précurseur ADP-glycero-manno-heptopyranose provenant de la voie de biosynthèse des lipopolysaccharides comme substrat pour modifier AIDA-I, dans un ratio d'environ 19 résidus heptose par molécule d'AIDA-I (7). L'heptose n'est pas un sucre retrouvé chez les eucaryotes et très peu d'exemples de modification d'une protéine bactérienne par ce dernier existent. Il est à noter que ce sucre précurseur est retrouvé dans le cytoplasme de la bactérie et n'est pas associé à un précurseur lipidique, comme c'est le cas pour les plupart des autres systèmes de glycosylation bactérien. Cette observation suggère donc que la glycosylation aurait lieu dans le cytoplasme de la bactérie, quoique qu'aucune preuve expérimentale ne soit disponible. De plus, à l'exception de la famille de N-glycosyltransférase représentée par HMW1C, aucune protéine glycosylée dans le cytoplasme d'une bactérie Gram négatif et utilisant le système général Sec n'a été décrite précédemment.

Une autre question relative à ce système concerne la reconnaissance des protéines cibles. En effet, la glycosylation par Aah et TibC est très spécifique. Les substrats de ces deux protéines (AIDA-I, TibA et Ag43) possèdent un motif imparfait de 19 acides aminés répété dans leur domaine extracellulaire (SGGxAxTx•NSGGxQx•x, où X représente n'importe quel acide aminé et • représente une valine, leucine ou isoleucine) (157). Comme ce motif est une caractéristique unique à ces protéines, il semblerait être le site idéal pour l'ajout spécifique de sucres. Basé sur l'observation que cette séquence répétée est particulièrement enrichie en glycine et sérine, il a été suggéré que l'ajout des sucres se fait sur les résidus

sérine, d'où la classification à titre de O-glycosyltransférase (7). Toutefois, aucune preuve expérimentale ne confirme cette hypothèse et il faut mentionner qu'on retrouve aussi dans ce motif un résidu asparagine qui pourrait être la cible de la glycosyltransférase. De plus, l'autotransporteur EhaJ est modifié par une glycosyltransférase homologue à Aah et TibC, mais ne possède pas le motif consensus de 19 acides aminés (228). Plusieurs questions restent donc d'actualité concernant ce groupe de glycosyltransférases, que ce soit au niveau de leur mécanisme d'action, de la reconnaissance des substrats ou de la fonction de cette modification particulière.

#### 2.2.4. Systèmes généraux de O-glycosylation

Récemment, il a été découvert que *N. gonorrhoeae* pouvait non seulement modifier les pilines du pili de type IV, mais aussi une variété de protéines, montrant pour la première fois un système général de O-glycosylation chez les bactéries Gram négatif (234, 235). Onze protéines de *N. gonorrhoeae* ont été décrites comme glycosylées. Ces protéines sont toutes transportées dans le périplasme et possèdent des domaines transmembranaires ou des caractéristiques propres aux lipoprotéines. Ces substrats sont responsables de fonctions diverses telles que le repliement protéique, la formation de ponts disulfures, des composants d'ABC transporteurs ou des protéines impliquées dans la respiration anaérobie (235). La nitrite réductase AniA est aussi une glycoprotéine de *Neisseria* retrouvée dans la membrane externe (234). La glycosylation de ces protéines dépend de la même voie que celle des pilines et les sucres ajoutés sont similaires (234). Pour l'instant, la relevance réelle de ce nouveau système de glycosylation général reste à déterminer. Une des hypothèses quant au rôle de la glycosylation est que la présence des sucres favoriserait une interaction fonctionnelle entre les protéines impliquées dans des fonctions complémentaires (principalement dans l'espace périplasmique).

Une autre question de taille par rapport à ce système est le mécanisme de reconnaissance du substrat. Pourquoi seule la piline et quelques autres protéines sont glycosylées chez *N. gonorrhoeae*? Aucun motif consensus n'a pu être déterminé par l'observation de la séquence primaire en acides aminés. Toutefois, à l'exception de la piline, les sucres semblent se retrouver dans des régions de faible complexité riches en résidus alanine, sérine et proline. Une surreprésentation du motif Ser-Ala-Pro-Ala a été observée, quoique ce motif

ne soit ni nécessaire ni suffisant pour expliquer la sélectivité de cette oligosaccharyltransférase.

Le second système général de O-glycosylation est retrouvé chez le genre *Bacteroides*. *Bacteroides fragilis* fait partie de la flore intestinale des mammifères. Une partie impressionnante du génome de ce micro-organisme est dédiée uniquement à l'acquisition, la dégradation et l'utilisation de polysaccharides provenant de son écosystème (236). Cette bactérie produit d'ailleurs plus de huit types d'exopolysaccharides de surface, dont quatre types contiennent du fucose provenant de l'hôte, en plus d'une variété de glycoprotéines (237). Cette particularité d'utiliser des molécules communes de l'hôte pour recouvrir sa surface est considérée comme du mimétisme moléculaire et permet de mieux comprendre la tolérance de l'hôte envers cette espèce prédominante de la flore intestinale. L'étude plus approfondie de ces glycoprotéines a permis de mettre en évidence un locus génique, similaire à ceux impliqués dans la biosynthèse des polysaccharides capsulaires, nommé *lgf* (*B. fragilis* glycosylation) (238). Cette région comprend, entre autre, une flippase putative et cinq glycosyltransférases. L'identification de ce locus suggère que, tout comme les autres systèmes généraux de glycosylation procaryotes, la synthèse des sucres se fait dans le cytoplasme avant d'être transféré dans le périplasme puis attaché de façon covalente sur les protéines cibles. Il a d'ailleurs été démontré que le transport dans le périplasme est essentiel pour la glycosylation (238).

Similairement à ce qui a été démontré chez *Neisseria*, les protéines glycosylées de *Bacteroides fragilis* sont localisées dans le périplasme ou exposées à la surface sous forme de lipoprotéines. Les fonctions associées à ces protéines sont principalement reliées à la physiologie, incluant des chaperonnes périplasmiques ainsi que des protéines impliquées dans la dégradation de peptides. Toutefois, ce système de glycosylation est essentiel pour la bactérie puisque la délétion du locus *lgf* entraîne un ralentissement de la croissance *in vitro* ainsi qu'une incapacité à entrer en compétition pour l'établissement d'une flore intestinale chez les mammifères (238).

Ce système de O-glycosylation est le seul pour lequel il a été possible d'identifier une séquence consensus pour la reconnaissance des protéines cibles. En effet, tous les résidus glycosylés (sérine ou thréonine) se retrouvent dans le motif suivant : Asp-Ser/Thr-Ala/Ile/Leu/Val/Met/Thr, où le dernier acide aminé doit posséder un groupement méthyle (238). Similairement, *Flavobacterium meningosepticum*, une bactérie appartenant aussi à



l'embranchement des *Bacteroidetes*, est capable de O-glycosyler certaines protéines sur des sites consensus similaires à ceux de *B. fragilis*, soit Asp-Ser ou Asp-Thr-Thr (239). Il est évident que ce motif consensus de seulement trois acides aminés se retrouve dans beaucoup plus de protéines que les huit identifiés dans l'étude initiale. Par analyse bioinformatique, il semble y avoir 1021 protéines potentiellement glycosylées chez *B. fragilis*, soit plus de la moitié des protéines extracytoplasmiques de cette bactérie (190). L'ajout de sucres sur 12 de ces nouvelles protéines, incluant quatre localisées au niveau de la membrane interne, a été démontré expérimentalement. Puisqu'à ce jour, toutes les protéines sécrétées possédant le motif consensus et ayant été testées pour la présence de sucres se sont révélées glycosylées, il ne serait pas surprenant que plusieurs centaines de protéines de *B. fragilis* soient effectivement modifiées. Ce constat renforce encore plus l'idée que la glycosylation protéique est essentielle pour la physiologie et la survie de ce micro-organisme.

#### 2.2.5. O-glycosylation chez les bactéries Gram positif

Chez les bactéries Gram positif, il existe aussi plusieurs exemples de protéines O-glycosylées. Le premier système de type général mis en évidence est celui de la O-mannosylation, particulièrement retrouvé chez les actinomycètes, un groupe de bactéries Gram positif possédant un fort contenu en G/C et qui inclut les genres *Mycobacterium*, *Corynebacterium* et *Streptomyces*. Plusieurs protéines O-glycosylées ont été identifiées chez *Mycobacterium* (240). Chez *Streptomyces coelicolor*, seule la protéine PstS a été décrite comme glycosylée, quoique les auteurs suggèrent que plusieurs autres protéines pourraient être potentiellement glycosylées (241). Ce système de glycosylation spécifique utilisant des résidus mannose, est similaire à celui du même nom retrouvé chez les eucaryotes (242). En utilisant *M. tuberculosis*, il a été démontré que ce type de glycosylation ne cible que les protéines membranaires ou sécrétées. De plus, seules les protéines sécrétées par le système de transport Sec peuvent être modifiées, ce qui suggère une interaction fonctionnelle entre le complexe de glycosylation et la machinerie de sécrétion (243).

Le mécanisme de glycosylation est assez similaire aux systèmes généraux décrits précédemment. Le sucre est synthétisé par un groupe d'enzymes cytoplasmiques en association avec un lipide membranaire, le polyprenol-phosphate. Une mannosyltransférase (Pmt) est responsable du transfert du sucre sur la protéine cible. Cette glycosyltransférase

membranaire est homologue à celle retrouvée chez les eucaryotes et possède un site actif qui fait face à l'extérieur, suggérant que la glycosylation a lieu suite à la sécrétion (243). À ce jour, la séquence exacte reconnue par l'O-mannosyltransférase n'est pas connue, quoique *a priori*, la glycosylation a lieu sur des résidus thréonine dans des régions particulièrement enrichies en thréonine et proline (244). Tout comme pour les eucaryotes, il a été suggéré la structure secondaire autour du site de glycosylation serait importante pour la reconnaissance par la glycosyltransférase (245).

Un deuxième système de O-glycosylation décrit chez les bactéries Gram positif est celui relié aux adhésines riches en sérine (serine-rich bacterial adhesins) (246). Plusieurs exemples de ce type d'adhésines glycosylées de haut poids moléculaire ont été décrits, incluant Fap1 de *Streptococcus parasanguinis* (247), GspB de *S. gordonii* (248), SrpA de *S. sanguinis* (249) et SraP de *Staphylococcus aureus* (250). De façon générale, la glycosylation est importante pour la maturation de ces protéines et, par conséquent, pour leur fonction d'adhésine. Ces protéines ont plusieurs caractéristiques communes dont une séquence signal exceptionnellement longue en N-terminal et un motif d'ancrage à la membrane en C-terminal. De plus, elles possèdent une région près de l'extrémité C-terminale contenant une répétition de la séquence S-E/V/I (plus de 1000 répétitions pour Fap1) suivie d'une région non répétée impliquée dans le phénotype d'adhésion (246). Finalement, toutes ces protéines riches en sérine font partie d'un opéron de gènes codant aussi pour un système de sécrétion Sec accessoire ainsi qu'une variété de glycosyltransférases. Contrairement à ce qui a été décrit précédemment, la glycosylation des ces protéines a lieu dans le cytoplasme de la bactérie (251). De plus, les sucres ne sont pas synthétisés sur un précurseur lipidique. Le sucre initiateur, le *N*-acetylglucosamine, est conservé pour toutes ces adhésines et sa présence est requise pour l'ajout séquentiel des autres sucres, parmi lesquels figure le glucose, le *N*-acetylgalactosamine et le rhamnose (252, 253).

La sécrétion de ces protéines hautement glycosylées est un point d'étude important. D'ailleurs, l'étude de ces protéines a permis de mettre en évidence un système de sécrétion Sec accessoire, qui inclus les protéines SecA2 et SecY2 (254). Il est à noter que des variations entre les espèces existent concernant la fonction de ces éléments de sécrétion. Toutefois, il est clair que le transport de ces protéines dépend d'un système qui leur est propre. L'étude de la protéine GspB a même démontrée qu'en plus de SecA2/SecY2,

d'autres gènes accessoires sont requis pour la sécrétion, soit *asp1* à *asp5*. Il faut mentionner que les protéines Asp4 et Asp5 présentent une forte similarité de séquence avec SecE (52% similarité) et SecG (55% similarité), respectivement. Ces observations suggèrent donc qu'un système Sec accessoire, différent de système Sec classique, est requis pour la sécrétion de ces adhésines. L'évolution d'un tel système Sec accessoire peut s'expliquer par la nature de la protéine à sécréter. En effet, pour permettre une glycosylation séquentielle complète avant le transport, il doit forcément y avoir un mécanisme permettant le couplage entre la glycosylation et la sécrétion. Une autre hypothèse émise est que la taille du canal formé par le système Sec accessoire serait plus importante, permettant ainsi le passage de ces protéines hautement glycosylées.

La présence de ce deuxième système Sec ouvre la porte à plein de questions comme la nature du signal permettant l'utilisation du système accessoire au lieu du classique. Un point important est que ces adhésines possèdent une séquence signal atypique, pouvant aller jusqu'à 90 acides aminés. Fait intéressant, cette caractéristique a aussi été observée pour les substrats de la famille de N-glycosyltransférases atypiques représentés par HMW1C, seul exemple de N-glycosylation cytoplasmique. La séquence signal de GspB est requise pour la reconnaissance par le système Sec accessoire et inhibe la sécrétion par le système Sec classique (255). Toutefois, des éléments autres que la séquence signal semblent aussi être requis pour le transport via le système Sec accessoire (256). Malgré la spécificité et le faible nombre de substrats utilisant cette voie alternative, l'étude de ce nouveau système ne pourra qu'améliorer notre connaissance du processus de sécrétion à travers la membrane plasmique.

### 2.3. Applications biotechnologiques

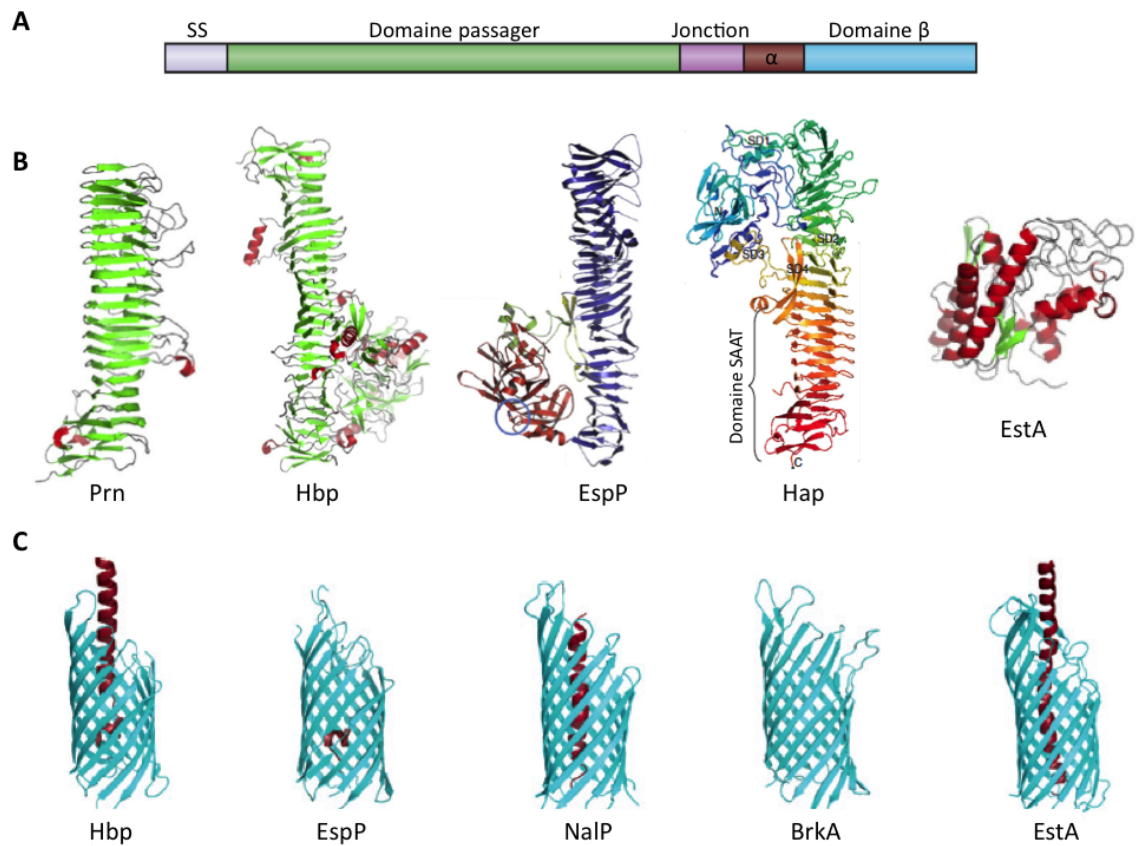
L'étude des différents systèmes de glycosylation bactériens, en plus d'être utile pour mieux comprendre la physiologie bactérienne ainsi que la biogenèse de nombreux facteurs de virulence, est aussi importante pour le développement de nouvelles applications biotechnologiques. Il est reconnu que beaucoup de glycoprotéines sont hautement immunogéniques. De plus, une réponse immunitaire de l'hôte dirigée contre des polysaccharides de surface, tels que les antigènes O et les polysaccharides capsulaires, est souvent efficace pour limiter ou éliminer la colonisation et l'infection causée par des

bactéries pathogènes (257). D'ailleurs, certains vaccins basés sur les polysaccharides capsulaires sont déjà utilisés en clinique (257).

Le transfert du système général de N-glycosylation de *C. jejuni* chez *E. coli* a été un élément déclencheur pour l'étude des applications biotechnologiques reliées à la glycosylation protéique chez les bactéries (197). En effet, la production *in vivo* de protéines recombinantes N-glycosylées chez une bactérie facilement manipulable et cultivable à peu de frais est un avantage majeur. Plus de 70% des protéines humaines thérapeutiques utilisées en clinique sont N-glycosylées (258). Les coûts de production de ces protéines sont souvent élevés du, en outre, à l'utilisation de lignées cellulaires spécifiques pour leur synthèse. Il a récemment été démontré qu'il était possible de générer, chez *E. coli* en modifiant la voie *pgl* de *C. jejuni*, des protéines eucaryotes N-glycosylées possédant le lien asparagine-*N*-acetylglucosamine normalement retrouvé chez les eucaryotes (259). De plus, le récent transfert fonctionnel simultané de quatre glycosyltransférases eucaryotes en conjonction avec l'enzyme PglB de *C. jejuni* chez *E. coli* ouvre la porte à de nouvelle façon d'exploiter la machinerie de glycosylation bactérienne (260). En résumé, la découverte de ces systèmes de glycosylation bactériens a ouvert de nouveaux champs de recherche qui vont, à coup sur, prendre plus d'importance au cours des prochaines années.

### **3. Autotransporteurs monomériques**

La famille des autotransporteurs monomériques appartient au système de sécrétion de type V, plus précisément au type Va (3). Ces protéines constituent le plus simple des systèmes de sécrétion chez les bactéries Gram négatif. Leur nom provient du fait que toute l'information nécessaire à leur sécrétion est présente dans un seul polypeptide. Toutefois, la biogenèse de ces protéines s'est relevée beaucoup plus complexe que présumée initialement et fait l'objet de nombreux débats. Dans les prochaines sections, les similitudes et différences structurelles présentes au sein de cette grande famille ainsi que la biogenèse de ces protéines seront discutées.



**Figure 7 : Organisation structurale des autotransporteurs monomériques** [figure adaptée des références (4, 261-263)]. **A.** Représentation schématique de l'organisation modulaire d'un autotransporteur monomérique classique (SS : séquence signal). **B.** Structures des domaines passagers provenant des autotransporteurs pertactin (Prn), Hbp, EspP, Hap et EstA. **C.** Structures correspondant au domaine  $\beta$  (tonneau  $\beta$  formé de 12 brins  $\beta$  en bleu et l'hélice  $\alpha$  en rouge) des autotransporteurs Hbp, EspP, NalP, BrkA et EstA.

### 3.1. Fonctions et structures des autotransporteurs monomériques

De façon générale, un autotransporteur monomérique est composé de quatre régions distinctes, soit une séquence signal N-terminale permettant la reconnaissance par le système Sec, un domaine passager fonctionnel, une région de jonction ou région auto-chaperonne et un domaine C-terminal (domaine  $\beta$ ) qui forme un tonneau  $\beta$  dans la membrane externe (figure 7A) (3).

### 3.1.1. Le domaine passager

Le domaine passager des autotransporteurs monomériques varie du point de vue de la longueur. En plus, la séquence en acides aminés est extrêmement variable et ne permet pas de relier une protéine avec la famille des autotransporteurs. Toutefois, avec la disponibilité de plusieurs structures, il est maintenant apparent que, malgré cette diversité de séquence primaire, les domaines passagers d'autotransporteurs sont construits sur un même thème structural, soit une conformation en hélice  $\beta$  droite (Figure 7B) (157). En fait, plus de 97% des domaines passagers sont prédits pour avoir une hélice  $\beta$  comme composant structural majeur (264).

La première structure à avoir été publiée est celle de la pertactin, une adhésine de *Bordetella pertussis* (265). Le domaine passager de la pertactin est formé d'une hélice  $\beta$  de 16 tours, où chaque tour d'hélice est composé de 3 petits brins  $\beta$  connectés entre eux par des boucles de longueur variable. Les structures d'autres autotransporteurs tels que l'hémoglobine protéase (Hbp) et la sérine protéase EspP retrouvées chez *E. coli* (261, 266) ainsi que l'immunoglobuline protéase A (IgAP) et l'adhésine Hap de *Haemophilus influenzae* (262, 267) ont montré une hélice  $\beta$  centrale, mais aussi différents domaines globulaires qui sortent de la structure hélicale de base. À titre d'exemple, le domaine passager de Hbp est d'une longueur plus importante que celui de la pertactin, formant ainsi une hélice  $\beta$  centrale de 24 tours (266). À l'extrémité N-terminale, on retrouve un domaine globulaire qui est responsable de la fonction sérine protéase de Hbp. Il semblerait donc que cette structure d'hélice  $\beta$  serve d'échafaud pour l'exposition de boucles ou domaines portant les résidus importants pour la fonctionnalité de la protéine. Cette architecture particulière pourrait aussi refléter une contrainte liée au mode de sécrétion de cette famille de protéines (voir plus bas). Il faut toutefois mentionner qu'il existe des exceptions, soit des domaines passagers d'autotransporteurs qui ne suivent pas ce thème structural commun. La structure entière de l'estérase EstA de *Pseudomonas aeruginosa* a récemment été résolue et montre une structure globulaire comprenant majoritairement des hélices  $\alpha$  et des boucles (Figure 7B) (263).

Une variété de fonctions a été attribuée aux domaines passagers d'autotransporteurs (3, 268). Beaucoup d'autotransporteurs sont aussi des protéines multifonctionnelles, signifiant que plusieurs fonctions peuvent leur être associées (269). Certains autotransporteurs ont été classés dans des sous-familles, basé en grande partie sur leur fonctionnalité. Par exemple,

chez les *Enterobacteriaceae*, on retrouve la famille de SPATEs (serine protease autotransporters of *Enterobacteriaceae*). Les protéines appartenant à cette famille ont toutes une activité principale de sérine protéase. Cette sous-famille inclue notamment Hbp, EspC, EspP, Vat, Sat, Pic, Tsh, Pet, EatA, EpeA, des autotransporteurs retrouvés chez différentes souches de *E. coli* causant des diarrhées ou de *E. coli* extra-intestinales comme les APEC (avian pathogenic *E. coli*) ou les UPEC (uropathogenic *E. coli*) (270). Un deuxième groupe d'autotransporteurs retrouvés chez *E. coli* est la sous-famille des SAATs (self-associating autotransporters) (5). Cette famille incluait originalement AIDA-I (Adhesin involved in diffuse adhérence), l'Ag43 (Antigène 43) et TibA (enterotoxigenic invasion locus B). Ces trois protéines sont impliquées dans l'adhésion des cellules épithéliales mais aussi dans l'agrégation bactérienne et la formation de biofilm, d'où le nom de cette sous-famille. Toutefois, cette classification en familles basée sur la fonction n'est pas parfaite. Par exemple, la protéine Hap de *H. influenzae* est impliquée dans l'adhésion aux cellules épithéliales des voies respiratoires ainsi qu'à diverses protéines de la matrice extracellulaire, dans l'agrégation entre bactéries via un domaine SAAT en plus de posséder un domaine sérine protéase similaire à celui des SPATEs (262). Les fonctions associées aux domaines passagers des autotransporteurs ne sont pas limitées à celles mentionnées précédemment. À titre d'exemple, il est possible de citer BrkA de *Bordetella pertussis* qui procure une résistance au sérum (271), IcsA de *Shigella flexneri* qui permet la polymérisation de l'actine et le mouvement intracellulaire (272), l'adhésine McaP de *Moraxella catarrhalis* qui agit comme estérase et phospholipase (273) et VacA de *H. pylori*, une toxine multifonctionnelle, qui induit la formation de vacuoles à l'intérieur des cellules cibles (274).

### 3.1.2. Le domaine $\beta$

La structure finale, soit lorsque intégrée dans la membrane externe, de cinq domaines  $\beta$  provenant de différentes espèces bactériennes sont maintenant disponibles (figure 7C) (263, 275-278). Ces cinq structures sont très similaires et se superposent facilement, suggérant ainsi que ce domaine est très conservé dans la famille des autotransporteurs. En effet, malgré l'absence de séquence primaire consensus, tous les domaines  $\beta$  semblent adopter une conformation en forme de tonneau  $\beta$  typique des protéines de membrane externe. Plus précisément, le domaine  $\beta$  forme un tonneau  $\beta$  comprenant 12 brins  $\beta$  reliés entre eux par

des boucles de longueur variable. L'intérieur du tonneau est de nature hydrophile et forme un pore aux dimensions internes de 10 x 12.5 Å (275). Une particularité commune à ce domaine  $\beta$  est la présence d'une hélice  $\alpha$  qui se localise au centre du tonneau. Cette hélice, présente chez tous les autotransporteurs monomériques est importante pour maintenir la forme du tonneau ainsi que pour fermer le pore afin de maintenir l'intégrité membranaire (279).

### 3.1.3. Le domaine de jonction

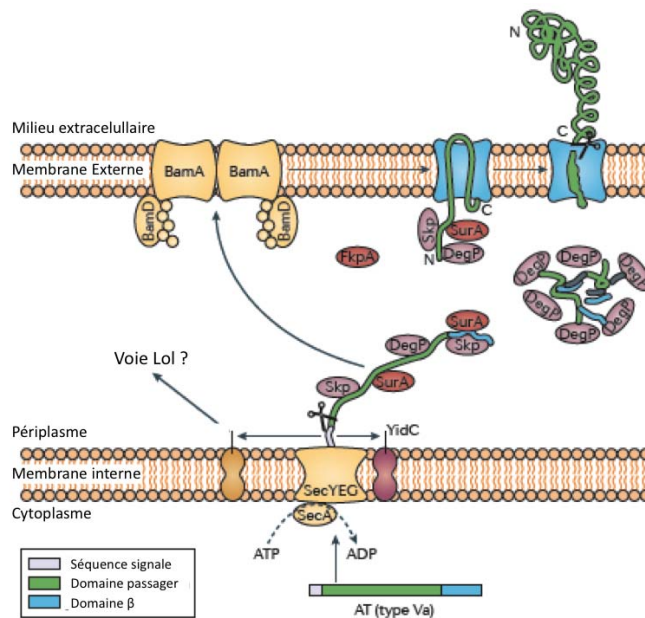
La région des autotransporteurs monomériques localisée entre l'hélice  $\alpha$  et le domaine passager est appelé domaine de jonction ou domaine auto-chaperonne. Cette région est aussi très bien conservée d'un point de vue structural (280). Plusieurs structures sont maintenant disponibles, la plupart d'entre elles ayant été cristallisées en même temps que le domaine passager. Une structure unique de la région de jonction est toutefois disponible pour l'autotransporteur IcsA de *S. flexneri* (281). La région de jonction possède un repliement similaire au domaine passager, soit une hélice  $\beta$  droite où chaque tour d'hélice est formé de trois brins  $\beta$ . Toutefois, la partie C-terminale adopte un repliement de type  $\beta$ -sandwich qui consiste en 2 brins  $\beta$  arrangés de façon perpendiculaire à l'hélice  $\beta$  centrale. Le rôle de cette région a été étudié chez plusieurs autotransporteurs, incluant BrkA, AIDA-I, Pet, IcsA, parmi d'autres (282-285). Une observation commune est que ce domaine est plus stable que le reste du domaine passager, c'est-à-dire qu'il est plus résistant à la protéolyse ou à la dénaturation thermique et chimique (281). C'est d'ailleurs la première région à être exposée lors du processus de translocation à travers la membrane externe (voir plus bas). Finalement, plusieurs études indiquent que cette région de jonction, qui adopte un repliement similaire à celui du domaine passager, servirait de modèle et de base pour le repliement subséquent du domaine passager (264, 286). Donc, la région de jonction a un rôle important pour la biogenèse des autotransporteurs, principalement au niveau du repliement extracellulaire.

## 3.2. Biogenèse des autotransporteurs monomériques

Les autotransporteurs monomériques sont des protéines exposées à la surface de la bactérie ou relâchées dans le milieu extracellulaire. Cela signifie que ces protéines doivent être



transportées dans le périplasma par un système de sécrétion de la membrane interne, traverser le périplasma en gardant une conformation adéquate, s'insérer dans la membrane externe et exposer le domaine passager à la surface et, dans certain cas, être clivées pour permettre la relâche du domaine passager dans le milieu extracellulaire. Leur nom autotransporteur provient du fait qu'aucun autre composant spécifique n'est requis pour leur sécrétion. Toutefois, comme il sera discuté dans la prochaine section, les protéines sécrétées via le système de sécrétion de type V utilisent les chaperonnes ainsi que les systèmes de sécrétion généraux pour permettre leur transport (Figure 8).



**Figure 8 : Biogenèse des autotransporteurs monomériques** [figure adaptée de la référence (4)]. Les autotransporteurs monomériques sont synthétisés dans le cytoplasme puis recrutés par le système Sec. Certains autotransporteurs possédant un motif lipoprotéine au niveau de la séquence signal vont être possiblement dirigés vers la voie Lol. La séquence signal de certains autotransporteurs semble interagir avec YidC lors du processus de sécrétion. Dans le périplasma, les autotransporteurs vont interagir avec des chaperonnes périplasmiques telles que DegP, SurA et Skp puis avec le complexe Bam, essentiel pour le repliement et l'insertion du domaine β. Le repliement du domaine passager se fait à la surface de la bactérie. Pour certains autotransporteurs, un événement de clivage a lieu après la sécrétion à travers la membrane externe, permettant la relâche du domaine passager dans le milieu extracellulaire.

### 3.2.1. Transport dans le périplasma

Comme pour toutes les protéines bactériennes, la transcription et la traduction a lieu dans le cytoplasme. La plupart des autotransporteurs possèdent une séquence signal classique pour le transport via le système général Sec, soit une séquence d'environ 20 à 30 acides aminés

et composée des régions n, h et c (3). Quelques protéines appartenant à ce groupe possèdent une séquence signal avec un motif typique des lipoprotéines (Leu-Ala/Ser-Gly/Ala-Cys) à l'extrémité C-terminale, qui permet la reconnaissance par le système Sec puis la prise en charge par la voie lol (localization of lipoproteins) pour une localisation finale au niveau de la membrane externe (100). Environ 10% des autotransporteurs monomériques possèdent une séquence signal anormalement longue d'environ 50 à 60 acides aminés et composée de cinq domaines distincts, nommés n1, h1, n2, h2 et c, où les domaines n2, h2 et c correspondent à un peptide signal classique (3). La séquence des domaines n1 et h1 est bien conservée et se nomme ESPR (extended signal peptide region). Quoique qu'il soit bien reconnu que les autotransporteurs monomériques, possédant ou non l'ESPR, utilisent le système Sec pour leur transport, la voie utilisée (co ou post-traductionnelle) pour se rendre au translocon SecYEG est toujours un sujet de débat. De plus, la présence d'une extension N-terminale à la séquence signal d'autotransporteurs appartenant à des espèces bactériennes éloignées d'un point de vue phylogénétique peut aussi suggérer une voie de sécrétion plutôt unique et inhabituelle pour ce groupe de protéines.

Comme mentionné, la fonction exacte de l'ESPR demeure un sujet de discussion. Les premières études sur le sujet ont montré que des séquences signal anormalement longues, soit celles de Hbp et EspP, deux autotransporteurs appartenant à la famille des SPATEs, permettaient un recrutement des polypeptides au translocon SecYEG par la voie SRP (67, 287). Cette observation est plutôt unique puisque le dogme sur la voie SRP est que seules les protéines intégrales de membrane interne sont sécrétées de façon co-traductionnelle. Toutefois, ces études ont été largement contredites par d'autres résultats obtenus avec Pet et EspP, deux SPATEs (288, 289). En effet, la présence de l'ESPR semble inhiber plutôt que promouvoir la liaison de SRP à la séquence signal, montrant que ces pré-protéines sont transportées dans le périplasme de façon post-traductionnelle. Finalement, une étude utilisant une technique de pontage chimique *in vivo* entre les pré-protéines et les composants de la machinerie de sécrétion Sec, montre qu'il y a une interaction entre les chaînes naissantes de Hbp et Ffh, un composant de SRP, ainsi qu'avec le TF, SecA et SecY (290). De façon générale, il est convenu que la plupart des autotransporteurs possédant ou non l'ESPR vont être dirigés vers le système Sec de façon post-traductionnelle par la voie SecB. Toutefois, certains des autotransporteurs monomériques possédant une séquence signal un peu plus hydrophobe, comme celle de Hbp, pourraient utiliser de façon facultative la voie SRP ou la voie SecB-dépendante (290).

Parallèlement, il a été suggéré que l'absence de l'ESPR chez EspP entraîne un mauvais repliement du domaine passager dans le périplasma, empêchant ainsi la sécrétion vers le milieu extracellulaire (291). Similairement, la délétion de la partie ESPR de l'autotransporteur Pet provoque une accélération du transport par le système Sec, suggérant ainsi que la présence de l'ESPR ralentit le processus de sécrétion (288). Il faut toutefois mentionner que la délétion de l'ESPR de Pet n'a pas d'effet sur la localisation ou la fonctionnalité de cette toxine, ce qui signifie que l'ESPR ne joue pas de rôle crucial dans le processus de biogenèse (292). Récemment, un nouveau modèle pour le rôle de l'ESPR a été suggéré, basé sur l'observation que les chaînes naissantes de Hbp pouvaient interagir avec YidC au niveau de la membrane interne (293). Plus intéressant encore, la déplétion de YidC entraîne une agrégation anormale de Hbp dans le périplasma, et par conséquent, une diminution du niveau de protéine exposé à la surface. Les auteurs proposent donc que YidC serait requis pour maintenir le polypeptide dans un état compatible avec le transport au niveau de la membrane externe. Cette observation a aussi été faite pour EspC, un autre membre de la famille des SPATEs (293). Il est important de mentionner que c'est le premier exemple d'une protéine transportée dans le périplasma utilisant YidC, généralement requise uniquement pour l'insertion des protéines de membrane interne. Toutefois, ce modèle reste à être vérifié pour d'autres autotransporteurs, tenant compte du fait que pour IcsA, le seul autotransporteur n'appartenant à la sous-famille de SPATEs à avoir été étudié du point de vue transport à travers la membrane interne, la déplétion de YidC n'a aucun effet (294).

D'un point de vue externe, ces résultats sont assez confus. En effet, la fonction exacte de l'ESPR dans la biogenèse des autotransporteurs demeure clairement indéterminée puisque pour chaque protéine étudiée, les résultats obtenus suggèrent un modèle de biogenèse différent, impliquant la voie SecB-dépendante uniquement ou SecB et SRP indifféremment. Toutefois, il est, selon moi, de plus en plus évident que l'ESPR de ces séquences signal permet le ralentissement du processus de sécrétion, avec des conséquences variables. Par contre, si ce ralentissement de la sécrétion est un facteur important pour la biogenèse, pourquoi seulement 10% des autotransporteurs monomériques possèdent cette extension? Un autre point à noter est que toutes ces études, à l'exception de celle portant sur IcsA de *S. flexneri*, ont été faites sur des autotransporteurs appartenant à la sous-famille des SPATEs. Pourtant, d'autres autotransporteurs ayant des fonctions différentes, tels que AIDA-I, TibA ou l'Ag43, possèdent aussi l'ESPR (5). Il serait donc intéressant de voir le rôle de cette

extension pour ces protéines ainsi que pour des autotransporteurs retrouvés chez des espèces bactériennes autres que *E. coli*.

### 3.2.2. Transit périplasmique

La plupart des protéines de membrane externe vont interagir avec des chaperonnes périplasmiques pour faciliter leur passage à travers le périplasme. SurA, une cis/trans peptidyl-prolyl isomérase, est la chaperonne périplasmique principale. Parallèlement, la voie Skp/DegP est aussi utilisée, surtout en condition de stress (87). Skp est une chaperonne générale tandis que DegP peut aussi bien agir à titre de chaperonne que de protéase, servant donc de système de contrôle qualité pour l'espace périplasmique. D'autres chaperonnes telles que FkpA, PpiA, ppiD, aussi des peptidyl prolyl isomérases, ainsi que les protéines Dsb, impliquées dans la formation des ponts disulfure, sont présentes dans le périplasme. Ces chaperonnes sont nécessaires pour maintenir les protéines de membrane externe dans une conformation compatible avec leur insertion/translocation au niveau de la membrane externe.

Comme la plupart des protéines de membrane externe, les autotransporteurs monomériques requièrent la présence de chaperonnes périplasmiques pour une sécrétion optimale, suggérant ainsi que la famille des autotransporteurs suit les mêmes règles générales que la plupart des protéines de membrane externe de *E. coli*. L'autotransporteur EspP de *E. coli*, est le mieux caractérisé du point de vue transit périplasmique. Une première étude a montré qu'il y a une interaction séquentielle entre EspP et les chaperonnes périplasmiques (295). En effet, le domaine passager de cette SPATE interagit initialement avec Skp puis avec SurA. D'autres études menées par le groupe de James P. Nataro ont confirmé ces résultats en plus de démontrer que le domaine passager de EspP pouvait aussi interagir avec DegP et FkpA (296, 297). Ces études ont aussi mis en évidence que l'interaction entre le domaine passager de EspP et les chaperonnes périplasmique était de nature transitoire, ce qui peut expliquer la variabilité des résultats obtenus. Dans le cas de Hbp, une interaction entre un intermédiaire de sécrétion, soit un mutant de Hbp bloqué au niveau de la membrane externe, et SurA a aussi été démontrée (298). Pour le domaine  $\beta$ , ces deux mêmes études ont démontrés soit une interaction avec Skp seul ou soit avec Skp et SurA (295, 296). Un des modèles suggérés pour EspP est donc que Skp serait la chaperonne principale qui interagirait avec le domaine  $\beta$  immédiatement après (ou possiblement pendant) le transport

par le système Sec, permettant ainsi au domaine  $\beta$  de rester dans une conformation adéquate (295). Skp et SurA entreraient en contact avec le domaine passager de façon transitoire et séquentielle, suivi d'un transfert au complexe Bam via SurA (voir la prochaine section).

Ces études visaient principalement à identifier des facteurs pouvant interagir avec les autotransporteurs lors de leur biogenèse. Toutefois, ces résultats ne tiennent pas compte de l'importance réelle de ces chaperonnes *in vivo*. Étudier la relevance de chaperonne générale comme SurA, Skp et DegP n'est pas facile, puisque leur inactivation a souvent un effet pléiotrope, du moins chez *E. coli*. Les premières études menées sur l'autotransporteur IcsA de *S. flexneri* ont démontré qu'une inactivation de Skp, DegP ou SurA a un impact négatif sur le phénotype associé à l'expression de IcsA, soit la motilité intracellulaire dépendante de la polymérisation de l'actine (299, 300). Toutefois, le niveau de tonneau  $\beta$  retrouvé dans la membrane externe est constant et similaire à la souche sauvage, en plus d'être correctement replié. Ces résultats suggèrent donc que l'inactivation d'une de ces chaperonnes n'entraîne pas de défaut d'insertion ou de repliement du tonneau  $\beta$ , mais plutôt du domaine passager. Des résultats concordants ont aussi été observés pour AIDA-I puisque le repliement *in vitro* du domaine  $\beta$  n'est pas influencé par la présence de Skp, SurA ou DegP (301). Il est aussi intéressant de noter que l'inactivation d'une chaperonne peut être complétée par la surexpression d'une autre, suggérant que l'interaction entre un autotransporteur et une chaperonne périplasmique n'est pas spécifique et exclusif (296, 300). L'ensemble de ces résultats montre que les protéines appartenant à la famille des autotransporteurs chez *E. coli* vont interagir avec des chaperonnes générales lors de leur passage dans le périplasma. Ce constat n'est pas surprenant puisque la plupart des protéines de membrane externe de *E. coli* requièrent aussi l'assistance de ces chaperonnes. Toutefois, le rôle exact de cette interaction, soit maintenir le tonneau  $\beta$  ou le domaine passager dans un état de repliement compatible ou tout simplement protéger le domaine passager contre une dégradation prématurée, reste à déterminer.

Il est intéressant de mentionner que ces études ont toutes été faites chez *E. coli* (ou *Shigella*). Quoique cette espèce reste un bon modèle d'étude, il est toujours instructif de regarder le mode de fonctionnement chez d'autres organismes. À titre d'exemple, chez *Bordetella pertussis*, SurA est une protéine essentielle. De plus, une nouvelle chaperonne de type parvuline, nommée Par27, interagit spécifiquement avec des protéines sécrétées via le système de type V, dont FHA, BrkA, Tcf et SphB1 (302). Quoique Par27 ne soit pas essentielle *in vivo* pour la sécrétion de ces substrats, l'existence d'une chaperonne

spécifique aux autotransporteurs ou aux systèmes de sécrétion à deux partenaires est intéressant. Chez *Neisseria meningitidis*, les chaperonnes périplasmiques ont une importance moindre que chez *E. coli*. D'ailleurs une inactivation de DegQ (homologue de DegP de *E. coli*), Skp ou SurA n'a pas d'effet sur l'insertion du tonneau  $\beta$  ou l'exposition extracellulaire du domaine passager pour l'autotransporteur IgA protéase ou pour l'autotransporteur trimérique NhhA (303).

### 3.2.3. Translocation au niveau de la membrane externe

Cette étape de la biogenèse des autotransporteurs est la moins bien comprise, principalement du fait qu'il a longtemps été supposé que les protéines de membrane externe s'inséraient de façon spontanée. Le complexe d'insertion de ces protéines, le système Bam, n'a été découvert que récemment et on commence seulement à comprendre son mode d'action. Globalement, cette étape de la biogenèse peut être regardée comme deux événements différents, soit l'insertion du tonneau  $\beta$  dans la membrane externe ainsi que la translocation et le repliement du domaine passager.

Il est maintenant clair que le complexe Bam est requis pour l'insertion du domaine  $\beta$  des autotransporteurs, au même titre que la plupart des protéines de membrane externe. Des études ont montré une interaction directe entre BamA et le domaine  $\beta$  de EspP et Hbp (298, 304). Par ailleurs, la déplétion de BamA a un effet négatif sur la biogenèse de plusieurs autotransporteurs monomériques, dont IcsA et AIDA-I (305). De plus, tous les autotransporteurs étudiés à ce jour possèdent le motif de reconnaissance C-terminal requis pour l'interaction avec le complexe Bam. Une étude réalisée avec EspP a démontré que le domaine  $\beta$  interagissait de façon séquentielle avec SurA, BamA, BamB puis BamD, suggérant ainsi un transfert du domaine  $\beta$  de la chaperonne périplasmique à BamA puis aux autres composants du complexe, possiblement impliqués dans les étapes plus tardives du processus d'insertion (295). Il est à noter que BamB, BamC et BamE ne semblent pas essentiels pour une insertion optimale des autotransporteurs monomériques (306).

L'étape de translocation du domaine passager au travers de la membrane externe est toujours un sujet d'étude qui mène à de nombreux débats. Deux modèles distincts sont présentement mis de l'avant, soit le modèle d'épingle à cheveux ou le modèle BamA. Le modèle BamA suggère que le complexe Bam, en plus d'être essentiel pour l'insertion du tonneau  $\beta$ , est aussi requis pour la sécrétion du domaine passager. La première hypothèse

est que BamA sert de pore pour la sécrétion du domaine passager à la place du domaine  $\beta$ . Cette hypothèse est peu plausible puisque cela suggère que les deux pores (BamA et le domaine  $\beta$ ) doivent s'ouvrir latéralement pour effectuer le transfert du domaine passager, ce qui implique un bris de nombreuses liaisons hydrogènes. De plus, il a récemment été montré que la séquence du domaine  $\beta$  ne peut pas être remplacée par celle d'une autre protéine de membrane externe, suggérant que le domaine  $\beta$  n'est pas seulement un pore mais joue aussi un rôle actif dans le processus de sécrétion (307). La deuxième idée suggérée est que le domaine  $\beta$  de l'autotransporteur s'insère en entier dans le pore de BamA, maintenant ainsi le tonneau dans un état compatible avec le passage du domaine passager (308). Toutefois, aucune preuve expérimentale ne permet de suggérer que BamA forme un pore d'une taille suffisante pour accommoder le tonneau  $\beta$  d'un autotransporteur monomérique. Alternativement, BamA et le domaine  $\beta$  pourraient simplement interagir lors du processus de sécrétion pour permettre au tonneau  $\beta$  de maintenir une conformation compatible avec le passage du domaine passager (298). Cette proposition reste toutefois hypothétique puisqu'une interaction entre BamA et un domaine passager n'a été observée que pour un intermédiaire de sécrétion bloqué au niveau de la membrane externe (298).

Le deuxième modèle général est celui de l'épingle à cheveux. Ce modèle suggère que la partie C-terminale de la protéine va être exposée à la surface en premier via le domaine  $\beta$ . Ceci implique que deux chaînes polypeptidiques doivent se trouver à l'intérieur du tonneau  $\beta$  simultanément, une chaîne resterait statique tandis que l'autre serait en mouvement pour permettre le passage du domaine passager entier du C-terminus vers le N-terminus. Plusieurs études ont présenté des résultats en accord avec ce modèle, principalement en utilisant des intermédiaires de sécrétion bloqués au niveau de la membrane externe (306, 307, 309, 310). Dans tous les cas rapportés, c'est la partie C-terminale de l'autotransporteur qui est exposée à la surface, tandis que la partie N-terminale du domaine passager se situe dans le périplasme de la bactérie. Le principal point de litige concernant ce modèle est la taille du pore. En effet, dans les structures de domaine  $\beta$  disponibles, le pore est relativement étroit et complètement occupé par une simple hélice  $\alpha$  (Figure 7C), ce qui suggère que seul deux chaînes polypeptidiques complètement dépliées peuvent être présentes en même temps dans le pore. Toutefois, plusieurs études ont démontré que des domaines passagers comprenant certains éléments de structure secondaire ou des boucles de petites tailles formées par des ponts disulfures pouvaient être sécrétées correctement (307, 311-313). Une réponse à cet argument est maintenant disponible grâce à la récente

structure de FimD, le composant localisé dans la membrane externe des Fimbriae de type I, avec ou sans substrat (314). FimD, un tonneau  $\beta$  de 24 brins, forme un pore dans la membrane externe qui permet le transport des sous-unités du fimbriae. La structure de FimD seule montre un pore de forme ovale (5.2 nm x 2.8 nm) tandis que cette même protéine en présence de FimH, une sous-unité du Fimbriae, a une forme plus circulaire (4.4 nm x 3.6 nm) permettant ainsi le passage de la protéine FimH complètement repliée. Cette étude est la première qui démontre que les tonneaux  $\beta$  sont des structures dynamiques qui peuvent modifier leur forme pour accommoder certains types de substrats. Cette observation ouvre donc la voie à de nombreuses spéculations, dont celle que les tonneaux  $\beta$  des autotransporteurs peuvent aussi subir un changement de conformation qui permettrait le passage de certains éléments de structures présent dans les domaines passagers. Ce modèle est donc celui qui est le plus plausible pour expliquer la translocation du domaine passager des autotransporteurs à travers la membrane externe.

Une des questions importantes à propos de la sécrétion du domaine passager à travers la membrane externe est la source d'énergie utilisée. En effet, la plupart des systèmes de sécrétion utilisent la force proton motrice ou des ATPases, qui peuvent être considérées comme des moteurs fournissant l'énergie mécanique requise afin de permettre le passage de molécules à travers une membrane biologique. Toutefois, le périplasma et la membrane externe sont dépourvues de sources d'énergie telles que l'ATP ou la force proton motrice. L'hypothèse principale est que le repliement du domaine passager lors de la sécrétion fournit l'énergie nécessaire (264). En C-terminal de la partie extracellulaire, se trouve le domaine de jonction, exposé à la surface en premier lors de la sécrétion. Cette région, plus stable que le reste du domaine passager, servirait de base pour le repliement (264, 286). Une étude récente faite avec la Pertactine et Pet démontre que la sécrétion des autotransporteurs est dépendante de l'anisotropie au niveau de la stabilité de la région sécrétée (région de jonction et domaine passager) (315). En fait, une région de jonction avec une grande stabilité ainsi qu'un domaine passager avec une faible stabilité permet une efficacité de sécrétion optimale. Alternativement, une région de jonction insuffisamment stable entraîne une diminution de la sécrétion à travers la membrane externe. Le modèle qui ressort de cette étude est donc que la sécrétion à travers la membrane serait en fait une diffusion directionnelle biaisée, basée sur les propriétés de repliement du domaine passager, expliquant donc pourquoi l'ATP n'est pas requis. Ce modèle concorde aussi avec une étude



précédente qui montrait que la mutation d'un acide aminé aromatique, situé au centre de l'hélice  $\beta$  formée par le domaine de jonction, empilé avec d'autres résidus aromatiques pour maintenir la conformation de l'hélice, cause une abolition de la sécrétion. Cette mutation influence très certainement la stabilité de la région de jonction, affectant ainsi la distribution de la stabilité chez cet autotransporteur (316). Il faut toutefois mentionner que pour l'autotransporteur BrkA, une délétion entière du domaine de jonction n'inhibe pas la sécrétion, mais résulte en un domaine passager moins stable et sensible à la dégradation protéolytique à la surface de la bactérie (282), ce qui suggère quand même que les événements de sécrétion et repliement peuvent être dissociés dans certaines circonstances.

Au moment du dépôt de cette thèse, une étude tout juste publiée a décrit un nouveau groupe de protéines retrouvé dans toutes les classes de protéobactéries et nommé TAM (translocation and assembly module), qui comprend TamA, une protéine intégrale de membrane externe appartenant à la superfamille BamA, et TamB, une protéine intégrale de membrane interne (317). Ces deux protéines forment un complexe trans-enveloppe qui est essentiel pour la sécrétion d'autotransporteurs monomériques à travers la membrane externe. Quoique intéressant, cet article du groupe de Trevor Lithgow génère plus de questions que de réponses. Comment le complexe TAM permet la sécrétion des autotransporteurs ? En interagissant avec les domaines passagers dans le périplasme ? Si oui, cela signifie que les autotransporteurs doivent interagir à la fois avec le complexe Bam (pour le domaine  $\beta$ ) et le complexe Tam, ce qui, théoriquement, nécessiterait un haut niveau de coordination. Est-ce que ce complexe est requis uniquement pour les autotransporteurs monomériques ou pour les trimériques et les membres des sous-familles Vd et Ve ? Un autre élément à considérer est qu'un complexe traversant l'enveloppe entière implique un accès aux sources énergétiques présentes dans le cytoplasme et remet en question tout ce qui a été assumé sur les autotransporteurs monomériques, soit un système de sécrétion qui ne requiert aucune source d'énergie et aucune protéine accessoire spécifique. Toutefois, à mon avis, d'autres études devront confirmer ces observations avant que le complexe TAM ne soit considéré comme un composant essentiel à la sécrétion par le système de sécrétion de type V.

### 3.2.4. Clivage

Plusieurs autotransporteurs monomériques sont clivés après la sécrétion à travers la membrane externe. Le clivage peut résulter en une relâche du domaine passager dans le milieu extracellulaire. Alternativement, dans certains cas, le domaine passager reste associé de façon non covalente à la surface de la bactérie. Les mécanismes de clivage décrits pour les autotransporteurs sont nombreux et variés. Certains autotransporteurs vont être clivés par d'autres protéines de membrane externe. À titre d'exemple, IcsA de *Shigella flexneri* est clivé par la protéase IcsP (318) alors que chez *N. meningitidis*, App, AusI, MspA et l'IgA protéase sont clivés par l'autotransporteur NalP (319-321). Toutefois, la plupart des autotransporteurs vont être clivés par un mécanisme autocatalytique. Dans le cas de Hap de *H. influenzae*, le clivage est de nature intermoléculaire où une molécule de Hap à la surface va couper une seconde molécule de Hap, via un domaine serine protéase (322). Pour les membres de la famille des SPATEs, le clivage est autocatalytique et de nature intramoléculaire (323). En effet, le clivage a lieu au niveau de l'hélice  $\alpha$  située à l'intérieur de tonneau  $\beta$  et implique un résidu aspartate et un résidu asparagine, tous deux bien conservés dans la famille des SPATEs. Ce mécanisme de clivage a aussi été démontré pour les autotransporter pertactin et BrkA de *Bordetella pertussis* (323). Il faut mentionner que le mécanisme de clivage pour plusieurs autotransporteurs tels que VacA (324), Ag43 (325), AIDA-I (326), les protéines polymorphiques de *Chlamydia pneumoniae* (327), parmi d'autres, demeure inconnu.

Quel est le rôle de ce clivage? La réponse est assez simple pour la plupart des autotransporteurs. Il est évident que les protéines qui ont des fonctions de nature cytotoxique, comme la toxine VacA de *Helicobacter pylori* (274), ou qui doivent être internalisées par les cellules hôtes pour avoir leur effet, comme pour l'entérotoxine Pet de *E. coli* (328), doivent être clivées et relâchées dans le milieu extracellulaire. IcsA est une protéine de *S. flexneri* localisée à un pôle de la bactérie et qui permet la polymérisation de l'actine et le mouvement intracellulaire (329). Le clivage de cette protéine, quoique partiel, est important pour maintenir la polarité et ainsi l'efficacité du déplacement bactérien (330). Dans le cas de l'adhésine Hap de *H. influenzae*, le clivage est aussi partiel et sert de mécanisme de régulation pour les fonctions d'adhésion et auto-agrégation (331). La fonction associée avec le clivage de certains autotransporteurs, particulièrement lorsque le domaine passager est clivé mais reste associé à la membrane externe comme pour AIDA-I

et Ag43 demeure inconnue. Ces deux protéines sont impliquées dans l'adhésion, l'agrégation bactérienne ainsi que la formation de biofilm, toutes des fonctions qui nécessitent la présence de la protéine à la surface de la bactérie (5).

### 3.2.5. Autres particularités

Certains autotransporteurs monomériques présentent des particularités qui leur sont propres et souvent requises pour leur fonction telles qu'une localisation polaire, l'ajout d'un groupement lipidique (lipoprotéine) ou l'ajout de sucres.

L'autotransporteur IcsA de *S. flexneri* est le modèle d'étude pour la localisation polaire. La présence de cette protéine à un seul pôle est clairement en lien avec sa fonction, soit permettre la polymérisation de l'actine et la formation d'une « queue » d'actine requise pour le mouvement intracellulaire (329). Toutefois, le mécanisme permettant cette localisation demeure encore inconnu. Il a été démontré que la localisation polaire s'effectuait dans le cytoplasme indépendamment de la sécrétion (294). De plus, IcsA est aussi localisée au pôle chez *E. coli*, *Salmonella typhimurium*, *Yersinia pseudotuberculosis* ou *Vibrio cholerae*, suggérant que le mécanisme de localisation est conservé chez plusieurs groupes de bactéries Gram négatif (332). Plusieurs études ont essayé de trouver un récepteur cytoplasmique de IcsA au pôle. DnaK, une chaperonne cytoplasmique ainsi que FtsQ, un composant de la machinerie de division cellulaire, sont tous les deux impliqués dans la localisation polaire de IcsA, quoique de façon indirecte (333, 334). D'autres autotransporteurs tels que SepA de *S. flexneri*, AIDA-I de *E. coli* et BrkA de *B. pertussis* sont aussi localisés au pôle de *Shigella* ou *E. coli*, quoique la présence de LPS complet est nécessaire (335). Le rôle de la localisation polaire de ces protéines reste toutefois inconnu et beaucoup moins évident que pour IcsA. En effet, SepA, un membre de la famille des SPATEs est sécrété dans le milieu extracellulaire et possède une activité sérine protéase (336). La localisation polaire ne joue probablement aucun rôle fonctionnel, si ce n'est que de permettre une concentration de protéines sécrétées au même endroit. BrkA est impliquée dans la résistance au sérum en interagissant directement avec des composants de la voie classique d'attaque du complément en plus d'avoir un rôle dans l'adhésion (337). Similairement, AIDA-I est impliqué dans l'adhésion aux cellules ainsi que dans l'agrégation bactérienne et la formation de biofilm (231, 338). Quel est donc l'avantage pour ces protéines d'être sécrétées à un pôle uniquement? Autre point intéressant, malgré

tous les efforts des dernières années et l'utilisation de plusieurs systèmes de criblage (333, 334), aucun récepteur direct permettant la localisation polaire n'a pu être identifié pour IcsA. Par conséquent, beaucoup de travail reste à faire pour comprendre la relevance et le mécanisme de localisation polaire de ces autotransporteurs.

Une autre modification retrouvée chez les autotransporteurs est l'ajout d'un groupement lipidique. En fait cette modification est plutôt unique et retrouvée chez seulement trois autotransporteurs, soit NalP de *N. meningitidis* (321), SphB1 de *B. pertussis* (164) et AlpA de *H. pylori* (339). NalP et SphB1 ont une bonne similarité de séquence et sont toutes deux responsables du clivage de d'autres protéines de membrane externe soit la protéine TpsA Fha pour SphB1 et plusieurs autotransporteurs pour NalP. Dans le cas de SphB1, l'ancrage lipidique est essentiel pour la fonctionnalité (340). La fonction de cet ajout pour NalP est moins bien connue. En effet, l'ajout d'un lipide ne permet pas la rétention du domaine passager de NalP à la surface, puisqu'un événement de clivage a lieu à environ 30 acides aminés de l'extrémité N-terminale de la protéine (321). Il a toutefois été suggéré que ce clivage de NalP pourrait être une façon de réguler son activité protéolytique envers d'autres autotransporteurs.

La dernière modification décrite chez les autotransporteurs est la glycosylation. L'ajout de sucres a seulement été décrit pour quatre autotransporteurs de *E. coli*, soit AIDA-I (7), TibA (8) et Ag43 (9) ainsi que EhaJ (228). Le mécanisme et la fonction de cette modification ont déjà été discutés dans une section précédente (voir section 2.2.3).

### 3.3. Utilisation biotechnologiques

Depuis la découverte de l'IgA protéase, plusieurs groupes ont vu le potentiel de ce nouveau type de système de sécrétion pour des applications biotechnologiques. La production de protéines recombinantes chez les bactéries, particulièrement chez *E. coli*, est hautement désirée du en partie aux plus faibles coûts d'utilisation ainsi qu'à la grande capacité de production. L'utilisation du système de sécrétion de type V pour la production de protéines recombinantes est avantageuse pour plusieurs raisons : C'est un système de sécrétion parmi les plus simple, plusieurs autotransporteurs sont naturellement produit chez *E. coli* et il permet soit la présentation d'une molécule à la surface, soit sa sécrétion dans le milieu extracellulaire, facilitant ainsi l'étape de purification (341). De plus, l'utilisation d'un hôte comme *E. coli* permet une manipulation génétique plus simple, telle que l'utilisation de

souches déficientes pour certains facteurs de repliement comme DsbA ou certaines protéases comme OmpT, optimisant ainsi l'utilisation d'autotransporteurs monomériques. AIDA-I est un des autotransporteurs les plus utilisés pour les applications biotechnologiques puisqu'il est présent chez *E. coli* et que sa biogenèse est relativement bien connue (342). Des exemples de présentation fonctionnelle de divers types d'enzymes, telles que des hydrolases, des estérases et des oxydoréductases, existent présentement dans la littérature.

Un tel système de présentation de peptides ou de protéines à la surface d'une bactérie peut servir pour de multiples applications. En effet, ce système peut servir pour la production basique de protéines hétérologues, comme outil de criblage pour la découverte de nouveaux produits pharmaceutiques ou commerciaux ou encore pour l'exposition d'épitopes ou antigènes à la surface pour le développement de vaccins vivants atténués. L'utilisation des autotransporteurs comme outil biotechnologique va grandement être facilitée par une meilleure connaissance de la biogenèse et de la structure de ces derniers, permettant éventuellement la présentation d'une variété de molécules à la surface des bactéries.

En plus de leur utilisation comme système de présentation, les autotransporteurs sont des cibles envisageables pour la génération de vaccin acellulaire. D'ailleurs, il existe présentement un vaccin acellulaire contre *B. pertussis* utilisant, entre autres, l'autotransporteur pertactin. Comme autre exemple, l'autotransporteur Hap de *H. influenzae* est aussi considéré comme une cible potentielle. En effet, le domaine passager de Hap permet le développement d'anticorps chez la souris, en plus de protéger des souris pré immunisées contre une colonisation des voies respiratoires (343).

#### **4. AIDA-I**

L'autotransporteur AIDA-I (Adhesin involved in diffuse adherence) retrouvé chez *E. coli* est un des autotransporteurs monomériques les plus étudiés, du point de vue de son rôle en tant que facteur de virulence, comme modèle d'étude pour la biogenèse de cette famille de protéines ainsi que comme outil de présentation de protéines hétérologues à la surface de *E. coli*. C'est aussi la protéine qui a servit de modèle d'étude pour la présente thèse de doctorat. La prochaine section est donc dédiée à la présentation spécifique de cet autotransporteur monomérique.

#### 4.1. Nomenclature

La classification principale des autotransporteurs monomériques a été faite par Henderson et collègues en 2004 (3). Dans cette dernière, on retrouve une famille d'autotransporteurs nommée AIDA. En fait, cette famille était considérée à ce moment comme la plus grande avec plus de 55 membres identifiés, incluant AIDA-I, Ag43 et TibA de *E. coli*, IcsA de *S. flexneri*, ShdA et MisL de *Salmonella enterica* ainsi que AutA et AutB de *N. meningitidis* (3). Cette classification était basée sur la présence du domaine COG3468, représentant les 644 acides aminés C-terminaux d'AIDA-I. Toutefois, on sait maintenant que cette région d'AIDA-I correspond principalement au domaine  $\beta$  et au domaine de jonction, expliquant ainsi le grand nombre de protéines appartenant à cette famille. En fait, les fonctions associées avec les membres de la famille AIDA-I sont variées et très peu de similarité au niveau de la séquence du domaine passager existe. Une étude plus récente a fait le recensement de tous les autotransporteurs présents dans les 28 génomes d'*E. coli*, complet ou partiel, disponibles en 2010 (344). Les auteurs ont identifiés, par analyse bioinformatique, 215 autotransporteurs appartenant à trois familles distinctes, dont les SPATES, les autotransporteurs trimériques et la famille AIDA. Tout comme la classification précédente, les membres de la famille AIDA sont reconnus uniquement par le domaine  $\beta$  et la région de jonction. Par contre, les auteurs ont poussé la caractérisation un peu plus loin en observant que tous les membres de cette famille possèdent des répétitions internes dans leur domaine passager, quoique la séquence et la longueur de ces répétitions varient énormément. De plus, le type de phénotype associé aux autotransporteurs de la famille AIDA semble assez bien conservé, incluant une implication dans l'adhésion et l'invasion de cellules de l'hôte, l'adhésion aux composants de la matrice extracellulaire, la formation d'agrégats bactériens ainsi que la formation de biofilm (344). Toutefois, comme cette étude reste uniquement centrée sur *E. coli*, on ne peut pas étendre ces observations aux autotransporteurs de type AIDA retrouvés chez d'autres espèces bactériennes. En résumé, la classification d'un autotransporteur dans la famille AIDA ne signifie pas nécessairement une similarité fonctionnelle ou au niveau de la séquence avec l'autotransporteur prototype retrouvé chez *E. coli* et nommé AIDA-I.

#### 4.2. Pathogénèse associée à l'expression d'AIDA-I

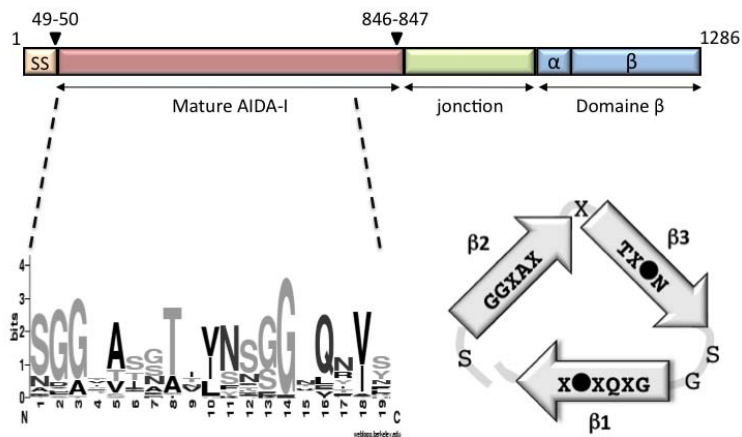
AIDA-I a été initialement identifiée dans un cas de diarrhée infantile causée par la souche de *E. coli* 2787 (O126 :H27) (338). Les auteurs ont rapidement démontré que l'expression d'AIDA-I était associée à une capacité des bactéries d'adhérer de façon diffuse à des cellules épithéliales, classant ainsi AIDA-I comme facteur de virulence retrouvé chez des souches de DAEC (diffuse-adhering *E. coli*) (338, 345). Toutefois, suite à la découverte d'AIDA-I, des études effectuées en France ont montré que très peu de souches de *E. coli*, soit entre 2 et 4 %, isolées de patients hospitalisés atteints de diarrhée, possédaient le gène codant pour AIDA-I (346, 347). En fait, des études plus récentes ont démontré qu'AIDA-I était plus souvent associé à des souches de *E. coli* causant des diarrhées post-sevrage et la maladie de l'œdème chez le porc. Ces études varient toutefois sur la prévalence d'AIDA-I dans les isolats cliniques, principalement du au pays où l'étude a été menée. Plus précisément, 41% des isolats en Allemagne (348), 63% en Belgique (349), 7% en Corée (350), 9% au Canada (351), 27% aux Etats-Unis (352) et 6,5% en Chine (353) se sont révélés AIDA positifs, suggérant un rôle assez important d'AIDA-I comme facteur de virulence chez le porc. Ces études ont aussi permis de montrer que la présence d'AIDA-I n'est pas limitée aux souches de DAEC, mais qu'AIDA-I est aussi retrouvé chez des isolats de type ETEC (enterotoxigenic *E. coli*) ou STEC (Shiga toxin-producing *E. coli*). Il est important de mentionner que dans toutes ces études, aucun isolat provenant de porcs sains possédaient le gène codant pour AIDA-I. De plus, une étude d'infection expérimentale chez des porcelets nouveaux-nés, utilisant la souche ETEC PD20 qui possède AIDA-I et la toxine STb comme facteurs de virulence, a démontré qu'AIDA-I est requis pour l'apparition des symptômes diarrhéiques (354).

Malgré l'implication principale d'AIDA-I dans l'adhésion chez le porc, très peu d'études, toutes effectuées en France, ont réellement testé la prévalence de cette protéine dans des isolats humains (346, 347). En fait, il semble y avoir une certaine similarité entre les séquences d'AIDA-I porcines et la séquence de la souche humaine 2787 (355). De plus, le gène codant pour AIDA-I est localisé sur un plasmide (338), ouvrant ainsi la possibilité d'un transfert entre souches de *E. coli* porcines et humaines. De façon globale, on considère qu'AIDA-I est requis pour l'attachement des bactéries aux parois intestinales des porcs, soit au niveau du mucus ou des entérocytes, permettant ensuite la libération de toxines dans un environnement restreint à proximité des cellules. Cette concentration de toxines va

permettre un effet maximal, soit une sécrétion d'électrolytes et une diminution de l'absorption d'eau par les cellules, entraînant ultimement une diarrhée (356).

#### 4.3. Structure et biogenèse d'AIDA-I

L'organisation structurale d'AIDA-I correspond à une organisation classique pour un autoporteur monomérique (Figure 9). AIDA-I est synthétisé sous la forme d'une pré-pro-protéine de 1286 acides aminés (357). Tout d'abord, cette protéine possède une séquence signal en N-terminal de 49 acides aminés avec l'ESPR caractéristique aux autoporteurs et permettant le transport à travers la membrane plasmique par le système Sec. Les événements menant la pré-pro-protéine au translocon n'ont jamais été étudiés directement pour AIDA-I et demeurent inconnus.



**Figure 9 : Organisation structurale d'AIDA-I.** En haut : représentation schématique de l'organisation modulaire d'AIDA-I. SS : séquence signal,  $\alpha$  : hélice  $\alpha$ . À gauche : Weblogo (<http://weblogo.berkeley.edu/>) représentant la conservation des acides aminés dans le motif répété de 19 acides aminés. À droite : Chaque répétition du motif consensus de 19 acides aminés est prédit pour former un tour d'hélice  $\beta$  en adoptant une structure secondaire composée de trois petits brins  $\beta$  (157).

Suite à l'enlèvement de la séquence signal, la pro-protéine est relâchée dans le périplasma de la bactérie. Pour le transport périplasmique d'AIDA-I, aucune donnée spécifique n'est disponible. Toutefois, une étude visant à caractériser le repliement *in vitro* du domaine  $\beta$  en association avec le domaine de jonction a montré que SurA, Skp et DegP ne pouvaient pas promouvoir le repliement de ce fragment d'AIDA-I en solution (301). De plus, par analyse de changement dans le profil de fluorescence intrinsèque, aucune interaction entre le domaine  $\beta$  d'AIDA et Skp n'a pu être observée. Toutefois, ces études ont été faites *in vitro*



sur une protéine tronquée. Sachant maintenant que, pour certains autotransporteurs, surtout ceux retrouvés chez *E. coli*, l'interaction avec des chaperonnes périplasmiques est requise pour une biogenèse optimale, une étude plus approfondie du transport périplasmique d'AIDA-I est à coup sur requise.

La partie C-terminale d'AIDA-I (acides aminés 952 à 1286) est formée d'un domaine  $\beta$  classique. Originellement, les programmes de prédiction de structure prévoyaient 14 brins  $\beta$  transmembranaires pour le domaine  $\beta$  (284). Toutefois, avec les cinq structures maintenant disponibles montrant toutes un domaine  $\beta$  d'autotransporteurs monomériques à 12 brins  $\beta$ , il est fort possible que ces observations s'appliquent aussi à AIDA-I. D'ailleurs, une modélisation plus récente, basée sur la structure de NalP, suggère qu'AIDA-I forme bien un tonneau  $\beta$  avec 12 brins (358). La séquence C-terminale d'AIDA-I contient le motif de reconnaissance pour le complexe Bam. Il a d'ailleurs été démontré que la déplétion de BamA affectait la biogenèse d'AIDA-I, suggérant, que comme la plupart des protéines de membrane externe, qu'AIDA-I utilise le complexe Bam pour l'insertion dans la membrane externe (305). Le domaine de jonction d'AIDA-I correspond aux résidus 847 à 951. Comme pour tous les autotransporteurs, cette région est essentielle pour la biogenèse de cette protéine (284). En fait, l'introduction de mutations à différents endroits entre le domaine de jonction et le tonneau  $\beta$  (positions 965, 974 et 976) permet la sécrétion du domaine passager à la surface mais dans une conformation anormale et sensible aux protéases, suggérant un rôle primordial pour le repliement extracellulaire du domaine passager (358).

Suite à la sécrétion du domaine passager dans le milieu extracellulaire, un événement de clivage a lieu. Le clivage a lieu entre les résidus Sérine 847 et Alanine 848 (326). Ce site de clivage fait la délimitation entre le domaine passager nommé AIDA-I mature (acides aminés 50 à 846) et le domaine translocateur AIDA<sub>c</sub> (acides aminés 847 à 1286) qui correspond à la région de jonction et au domaine  $\beta$  (Figure 9). Ce clivage ne mène toutefois pas à la relâche du domaine passager dans le milieu extracellulaire. AIDA-I mature reste associé à la membrane externe par un mécanisme inconnu, quoique d'un traitement à la chaleur permet sa relâche, suggérant fortement une interaction de type non-covalente. Il a été suggéré que le clivage d'AIDA-I est de nature auto-catalytique puisqu'il est indépendant des principales protéases extracytoplasmiques telles que OmpT, OmpP et DegP (326). De plus, le clivage a lieu chez d'autres enterobacteriaceae, notamment chez *Salmonella* et *Shigella*, suggérant qu'AIDA-I n'est pas clivé par un facteur spécifique de *E.*

*coli*. Malgré ces évidences, aucun mécanisme de clivage autocatalytique n'a pu être déterminé et les mécanismes déjà connus ne peuvent s'appliquer à AIDA-I (322, 323). En effet, AIDA-I ne possède pas de domaine sérine protéase comme l'autotransporteur Hap de *H. influenzae*, pas plus que les résidus aspartate et asparagine essentiels au clivage des SPATEs. De plus, le rôle de ce clivage extracellulaire n'a jamais été étudié.

D'un point de vue structural, la partie AIDA-I mature est principalement enrichie de brins  $\beta$ , tel que déterminé par dichroïsme circulaire (359). En fait, selon les programmes de prédiction, AIDA-I mature posséderait une structure typique des autotransporteurs monomériques, soit une hélice  $\beta$  similaire à celle de la pertactin de *B. pertussis* (265) (voir Figure 7B). Une des caractéristiques d'AIDA-I mentionnée précédemment est la présence de répétitions dans le domaine passager. AIDA-I possède, à partir de l'extrémité N-terminale, 35 répétitions du motif imparfait de 19 acides aminés SGGxAxTx•NSGGxQx•x, où x représente n'importe quel acide aminé et • représente une valine, leucine ou isoleucine (157) (Figure 9). Ce motif répété serait la base même de l'hélice  $\beta$  prédite pour le domaine passager, où chaque répétition permet la formation de trois petits brins  $\beta$  constituant un tour d'hélice (157).

Le dernier point concernant la biogenèse d'AIDA-I est la glycosylation. Comme mentionné précédemment (voir section 2.2.3), AIDA-I est glycosylé par une heptosyltransférase spécifique nommé Aah (7). *A priori*, la glycosylation a lieu sur le motif répété d'AIDA-I, quoique aucune preuve expérimentale ne supporte cette hypothèse. De plus, le rôle de la glycosylation d'AIDA-I demeure incertain.

#### 4.4. Fonctions associées à l'expression d'AIDA-I

Comme mentionné précédemment, AIDA-I a été initialement découvert pour son rôle dans l'adhésion aux cellules (338). Depuis, différentes études ont essayé de trouver un récepteur spécifique pour AIDA-I, sans succès réel. Une glycoprotéine membranaire (gp119) de 119 kDa a été identifiée à partir de cellules HeLa (359). Toutefois, les auteurs n'ont pas réussi à déterminer l'identité de cette gp119. Par contre, ils ont mis en évidence que AIDA-I permet l'adhésion à une variété de lignées cellulaires provenant de différents tissus et espèces, suggérant une distribution plutôt ubiquitaire du récepteur (359). Un second groupe a recherché un récepteur dans le mucus présent à la surface des cellules (360). Une protéine cytoplasmique a été identifiée pour son affinité avec AIDA-I, mais son implication réelle

dans l'adhésion n'a pas été démontrée. Finalement, AIDA-I ne semble pas être impliqué dans l'adhésion aux protéines de la matrice extracellulaire (229). Une implication d'AIDA-I a aussi été démontrée dans la capacité des bactéries à auto-agrégéer ainsi qu'à former des biofilms (231). La capacité d'AIDA-I de permettre l'agrégation bactérienne provient de sa capacité d'interagir avec elle-même avec une haute affinité (361).

AIDA-I fait partie de la famille des SAATs, qui inclus aussi l'Ag43 et TibA (5). Ces trois autotransporteurs glycosylés présents chez *E. coli* sont impliqués dans des phénotypes similaires et possèdent des homologies structurales ainsi qu'au niveau de la séquence primaire, dont le motif répété de 19 acides aminés. Toutefois, le rôle de Ag43 dans l'adhésion n'est pas clairement défini et TibA permet aussi l'invasion des cellules en cultures. De plus, la longueur de ces trois protéines est différente et TibA possède un domaine supplémentaire, soit une région riche en proline située entre le domaine de jonction et le domaine  $\beta$ . Finalement, l'Ag43 est une protéine retrouvée autant dans des souches pathogènes que commensales de *E. coli*, contrairement à AIDA-I et TibA qui sont, pour l'instant, uniquement associées à des souches causant des diarrhées. Donc, malgré leur similarité, quelques différences persistent au sein de cette sous-famille d'autotransporteurs monomériques.

Malgré nos nombreuses connaissances et les avancées de ces dernières années sur la biogenèse et la fonction des autotransporteurs monomériques, de nombreuses questions restent en suspens. Pour les travaux de cette thèse, l'autotransporteur AIDA-I a été utilisé comme modèle pour étudier la fonction des autotransporteurs appartenant à la famille des SAATs. De plus, plusieurs aspects de la biogenèse d'AIDA-I sont inconnus, tant au niveau du mécanisme de sécrétion qu'au niveau des modifications particulières à AIDA-I, comme le clivage protéolytique ou la glycosylation. Globalement, voici quelques exemples de questions concernant AIDA-I dont les réponses demeurent inconnues et qui ont guidé les travaux présentés dans cette thèse : Quel est le rôle de l'ESPR pour la biogenèse d'AIDA-I, toutes les études précédentes ayant été uniquement faites avec des membres de la famille des SPATEs ? Comment AIDA-I est dirigé au système de sécrétion Sec ? Comment AIDA-I est clivé ? Quelle est la fonction associée au clivage de AIDA-I ? Est-ce qu'il y a des conditions qui permettent la relâche d'AIDA-I mature dans le milieu extracellulaire ? À quoi sert la glycosylation d'AIDA-I ? Dans quel compartiment cellulaire a lieu la glycosylation ? Comment fonctionne Aah, la glycosyltransférase associée à AIDA-I ? Est-

ce que la glycosylation de AIDA-I est unique ou cette modification est répandue dans la famille des autotransporteurs ? Est-ce que la fonctionnalité de AIDA-I se limite aux fonctions déjà connues ?

Les travaux présentés dans cette thèse apportent une réponse à la plupart des questions précédentes et fournissent par le fait même de nouvelles informations permettant de mieux comprendre la biogenèse d'une des plus importantes familles de facteurs de virulence bactériens.

## ARTICLE 1

**Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I)**

Journal of Bacteriology, 2006, Dec., Volume 188 (24), pages 8504-12

## Contribution de l'auteur pour cet article

En tant que premier auteur de cet article, j'ai effectué la plupart des expérimentations décrites, à l'exception d'une construction génétique, pAg, qui permet l'expression de AIDA-I glycosylé. J'ai aussi écrit la totalité de l'article.

## Contribution scientifique de l'article

Dans cet article, nous avons décrit pour la première fois une procédure permettant la purification de AIDA-I sous sa forme native, à partir de la membrane externe. Cette technique de purification nous a permis de constater que les deux fragments issus du clivage, soit AIDA-I mature et l'unité de translocation AIDAc, peuvent être co-purifiés et restent fortement associés l'un à l'autre, mettant ainsi en évidence le mécanisme d'association de AIDA-I avec la membrane externe. De plus, nous avons effectué différentes mutations ponctuelles au niveau du site de clivage afin de bloquer ce dernier. Par l'étude d'un variant pour lequel le clivage est pratiquement aboli, nous avons démontré que dans les conditions *in vitro* utilisées, le clivage n'est pas requis pour la fonctionnalité ou la stabilité de AIDA-I. Finalement, nous avons démontré que l'expression de AIDA-I permet aussi l'invasion de cellules en culture, au même titre que l'invasine TibA, un autre membre de la famille des SAATs.

**Proteolytic Processing is not Essential for Multiple Functions of the *Escherichia coli*  
Autotransporter Adhesin Involved in Diffuse Adherence (AIDA-I)**

Running title: Cleavage of the AIDA-I autotransporter

Marie-Ève Charbonneau, Frédéric Berthiaume and Michael Mourez\*

Canada Research Chair on Bacterial Animal Diseases, Université de Montréal, Faculté de  
Médecine Vétérinaire, 3200 Sicotte, St-Hyacinthe, J2S 7C6, Québec, Canada

**ABSTRACT**

The *Escherichia coli* Adhesin Involved in Diffused Adherence (AIDA-I), like many other autotransporter proteins, is released in the periplasm as a proprotein undergoing a proteolytic processing after its translocation across the outer membrane. The proprotein is cleaved into a membrane-embedded fragment, AIDAc, and an extracellular fragment, the mature AIDA-I adhesin. The latter remains non-covalently associated with the outer membrane and can be released by heat treatment. The mechanism of cleavage of the proprotein and its role in the functionality of AIDA-I are not understood. Here, we show that cleavage is independent of the amount of AIDA-I in the outer membrane, suggesting an intramolecular autoproteolytic mechanism or a cleavage mediated by an unknown protease. We show that the two fragments, mature AIDA-I and AIDAc, can be co-solubilized and co-purified in a folded and active conformation. We observed that the release by heat treatment results from the unfolding of AIDA-I and that the interaction of AIDA-I with AIDAc seems to only be disturbed by denaturation. We constructed an uncleavable point mutant of AIDA-I, where a serine of the cleavage site was changed into a leucine, and showed that adhesion, auto-aggregation and biofilm formation mediated by the mutant are indistinguishable from the wild-type levels. Lastly, we show that both proteins can mediate the invasion of cultured epithelial cells. Taken together, our experiments suggest that the proteolytic processing of AIDA-I plays a minor role in the functionality of this protein.

## INTRODUCTION

Adherence is a crucial step in the development of infectious diseases. One group of diarrheagenic strains of *Escherichia coli* is characterized by their pattern of diffuse adherence (DA) on the surface of epithelial cells and therefore called diffusely adhering *E. coli* (DAEC). One afimbrial adhesin was shown to mediate DA and called Adhesin Involved in Diffuse Adherence (AIDA-I) (2). AIDA-I was originally cloned as a plasmid-encoded protein from a DAEC strain isolated in a case of infantile diarrhea. This adhesin has since been shown to play a role in the pathogenesis of *E. coli* strains causing post weaning diarrhea or edema disease in piglets, two infections that cause major economic losses in farms worldwide (32, 38, 39). AIDA-I has been shown to be one of the adhesins mediating DA, along with the fimbrial adhesin F1845 and others (5). In addition to its adhesive properties, AIDA-I has also been implicated in biofilm formation and auto-aggregation of bacterial cells (50). The *aidA* gene that codes for the autotransporter adhesin is associated with a second gene, named *aah* for autotransporter adhesin heptosyltransferase, which codes for a cytoplasmic glycosyltransferase (3).

AIDA-I is secreted by the monomeric autotransporter pathway of the type V secretion system (21). To date, most autotransporters identified are implicated in virulence (20), and many, like AIDA-I, are involved in adhesion (17). AIDA-I, a protein of 140 kDa, is produced as a pre-proprotein, processed at both N-terminal and C-terminal ends (Fig. 1). Its translocation occurs in three steps: First, the pre-protein is secreted across the inner membrane, presumably via the *sec* secretion system, as observed with other autotransporters (6, 46). In this process, an N-terminal signal sequence of 49 amino acids is cleaved, releasing the proprotein in the periplasm. The second step is the insertion of a C-terminal domain of the proprotein in the outer membrane. Lastly, the rest of the proprotein is translocated through the outer membrane, presumably via the membrane-inserted C-terminal domain (1, 53). After the translocation at the cell surface, the proprotein is cleaved into a membrane-embedded domain of 45 kDa, AIDAc, and a 100 kDa domain exposed to the extracellular milieu, the mature AIDA-I adhesin (Fig. 1). This cleavage is not mediated by the known periplasmic or outer membrane proteases (DegP, OmpT or OmpP) and occurs in *E. coli* as well as in *Salmonella* or *Shigella* strains, suggesting that it is an autocatalytic event (53). However, no protease catalytic site has been identified in the protein.



After its outer membrane cleavage, the mature AIDA-I remains non-covalently associated to the cell surface and can be released after a brief heating at 60°C. To date, no role has been attributed to this cleavage. Other autotransporters have been shown to be cleaved. The *Shigella flexneri* autotransporter protein VirG mediating intracellular spreading through actin polymerization is cleaved by a specific outer membrane protease, SopA/IcsP (10). The cleavage has been shown to be dispensable for its function but could be involved in the polar localization of this protein on the bacterial cell surface (8, 10, 16). The *Haemophilus influenzae* autotransporter adhesin Hap is also cleaved after translocation (52). Inhibition of the autoproteolytic cleavage of Hap increased bacterial auto-aggregation and adhesion, suggesting that the cleavage acts as a regulation mechanism (15, 23, 24). These proteins, however, are significantly different from AIDA-I, in sequence and function, thus the mechanism and function of the cleavage of AIDA-I remain obscure.

TibA, another adhesin autotransporter cloned from an enterotoxigenic strain of *E. coli*, is similar to AIDA-I both in sequence and function: Both proteins have almost identical repetitive amino acid sequences in their extracellular domain (12), and both can mediate biofilm formation, bacterial auto-aggregation and cell adhesion (51). Both are modified by the addition of heptose residues, and their corresponding glycosyltransferases can be functionally exchanged (30, 35). Two major differences exist between these two proteins, however: TibA is not cleaved after secretion and it mediates the invasion of epithelial cells. It is therefore tempting to speculate that the outer membrane cleavage of AIDA-I could be one reason for the functional difference between AIDA-I and TibA. Interestingly, AIDA-I and TibA have recently been grouped with the *E. coli* autotransporter Ag43 (43) in a new class of virulence factors tentatively named self associating autotransporters or SAAT (28). Like TibA and AIDA-I, Ag43 is glycosylated and mediates adhesion to cultured epithelial cells (49), has a similar 19 amino acid repeats and promotes autoaggregation (44). Furthermore, like AIDA-I, Ag43 is cleaved by an unknown mechanism after its outer membrane translocation and is not known to mediate invasion (7).

In this study, we characterized the outer membrane cleavage of AIDA-I. We observed that this cleavage occurs at a rate independent of the level of expression of AIDA-I and that it is dramatically reduced by a single point mutation at the cleavage site. When AIDA-I is cleaved, we observed that the mature AIDA-I is strongly associated to AIDAc, as the two polypeptides can be co-solubilized by detergents and co-purified in a native

conformation. Lastly, using the uncleaved mutant of AIDA-I, we showed that cleavage at the cell surface is not essential for the stability of AIDA-I or its function, including the ability to invade an epithelial cell line, an activity which was not previously described.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** The *Escherichia coli* K12 strains C600 (F<sup>-</sup>, *thr1*, *leuB6*, *thi1*, *lacY1*, *supE44*, *rfdD1*, *fhuA21*) and UT5600 (F<sup>-</sup>, *ara14*, *leuB6*, *secA6*, *lacY1*, *proC14*, *tsx67*,  $\Delta$ (*ompT-fepC*)266, *entA403*, *trpE38*, *rfdD1*, *rpsL109*, *xyl5*, *mtl1*, *thi1*), obtained from New England Biolabs, were used in this study. We amplified by PCR the *aah* gene from the *E. coli* strain 2787 (2), kindly provided by John M. Fairbrother, Université de Montréal, using primers Aah-1F and Aah-1R (Table 1) and cloned the resulting fragment in the empty vector pTRC99a (Pharmacia Biotech) to produce plasmid pTRC-aah. We then amplified by PCR the whole operon including the *aah* and the *aida* genes from the *E. coli* strain 2787 using the expand long template PCR system (Roche) with primers A-1F and A-1R (Table 1). The amplified fragment was then digested with *Xba*I and *Age*I and cloned at the same sites in the pTRC-aah plasmid, generating the construction pAg, containing the whole *aah-aida* operon. The S846L, S846G, S846C, A847R, A847I and A847T mutations were introduced in pAg by oligonucleotide-directed mutagenesis performed with the QuikChange II site-directed mutagenesis kit (Stratagene) using the primers listed in Table 1 and their corresponding complementary oligonucleotides. A six histidines and glycine tag (HisG) was introduced in the N-terminal part of AIDA-I in constructions pAg and pAgS846L, the construct allowing expression of AIDA-I bearing the S846L mutation, by site directed mutagenesis using the primer listed in Table 1 and its complementary oligonucleotide, generating plasmids pAgH and pAgHS846L respectively. All constructions were verified by restriction mapping and sequencing.

**Bacterial and cell culture growth conditions.** Bacteria containing the different plasmids were grown on LB agar plate or in liquid LB medium containing 100  $\mu\text{g}\cdot\text{ml}^{-1}$  of ampicillin. Bacterial cultures were grown at 30°C. Growth was followed by measuring turbidity as an optical density at 600nm (OD<sub>600nm</sub>). At an OD<sub>600nm</sub> of 0.8 the cultures were induced overnight with 10  $\mu\text{M}$  of isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG), unless indicated otherwise. This low concentration of IPTG was used to limit the toxicity associated with the overexpression of AIDA-I.

Hep-2 cells (ATCC CCL-23) were grown at 37°C with 5% CO<sub>2</sub> in Dulbecco's modified Eagle medium (DMEM, Gibco) containing 10 mM sodium pyruvate (Sigma), bovine growth serum (Hyclone), 2.5 µg.ml<sup>-1</sup> of fungizone and 100 µg.ml<sup>-1</sup> of penicillin/streptomycin (Gibco).

**SDS-PAGE and immunoblotting.** Protein samples were diluted in twice-concentrated SDS-PAGE loading buffer containing β-mercaptoethanol, and denatured by heating at 100°C for 10 min. The samples were then separated by SDS-PAGE on 10% acrylamide gels. The gels were either stained with Coomassie blue or transferred to polyvinylidene fluoride membranes (Millipore). Immunodetection was performed with a serum raised against the heat-extracted mature AIDA-I (a generous gift of Dr M. Ngeleka, University of Saskatchewan), diluted 1:10,000 in blocking buffer (5% skim milk, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Triton X-100). A goat anti-rabbit horseradish peroxidase (HRP) conjugated antibody (Sigma) was used as a secondary antibody, according to the instructions of the manufacturer. Alternatively, an anti-HisG HRP-coupled antibody (Invitrogen) was used diluted 1:5,000 in the blocking buffer for the detection of proteins containing the HisG tag. Immune complexes were revealed using a 3,3',5,5'-Tetramethylbenzidine solution (Sigma).

**Preparation of membrane fractions.** Subcellular fractionation was performed essentially as previously described (48). Briefly, bacterial cultures, normalized at the same OD<sub>600nm</sub>, were harvested and resuspended in 950 µl of 10 mM Tris-HCl pH 8, 0.75 mM sucrose. Lysozyme (0.4 mg.ml<sup>-1</sup> final) and EDTA pH 8 (10 mM final) was added and the sample kept on ice for 30 min. The resulting spheroplasts were collected and lysed in Tris-buffered saline (TBS, 50 mM Tris-HCl pH 8, 150 mM NaCl) containing proteases inhibitor (Complete mini, Roche) using an ultrasonic processor. The lysates were centrifuged for 1 h in an ultracentrifuge at 250,000 x g. The pellets, containing bacterial membranes, were resuspended in TBS, and submitted to SDS-PAGE and immunoblotting, as described above.

**Analysis of AIDA-I outer membrane processing.** *E. coli* C600 harboring plasmid pTRC99a or pAg were grown until an OD<sub>600nm</sub> of 0.4 and induced for 3 h with 10 µM or 50 µM of IPTG. Normalized cultures were centrifuged for 10 min at 12,000 x g in microfuge tubes and the pellets were resuspended in 100 µl of phosphate-buffered saline (PBS).

Alternatively, the bacterial cultures were grown to an  $OD_{600nm}$  of 0.4 and induced with 10  $\mu$ M of IPTG. Aliquots (200  $\mu$ l) were taken every 10 min after induction for the following 2 h. The samples were centrifuged for 10 min at 12,000 x g in microfuge tubes and the pellets resuspended in 20  $\mu$ l of PBS. All samples were processed by SDS-PAGE and immunoblotting, as described above.

**Extraction of the cleaved mature AIDA-I.** Overnight cultures from *E. coli* C600 or UT5600 harboring empty vector pTRC99a, pAg, or pAgS846L were normalized at the same  $OD_{600nm}$  in 25 ml of LB broth. Bacteria were harvested and resuspended in 375  $\mu$ l of 10 mM sodium phosphate buffer, pH 7. In order to release the cleaved mature AIDA-I, the samples were heated at various temperatures between 30°C and 60°C for 20 min or treated for one hour with SDS 0.1%, Triton X-100 1%, Tween-20 0.01%, 10 mM Glycine-HCl pH 2, 100 mM Sodium Carbonate pH 11, or urea 0.1M and 2M (final concentrations). The treated samples were centrifuged for 5 min at 12,000 x g in microfuge tubes. A 10  $\mu$ l aliquot from the supernatant was submitted to SDS-PAGE and stained with Coomassie blue, as described above.

**Purification of AIDA-I.** Heat extracts or membrane fractions were prepared as described above, except that one liter of overnight cultures of *E. coli* C600 harboring pAgH or pAgHS846L were used and all volumes were modified accordingly. The membranes were resuspended in TBS containing 1% Triton X-100, incubated for one hour and centrifuged for 1 h in an ultracentrifuge at 250,000 x g. The histidine-tagged AIDA-I contained in the solubilized membranes or the heat extracts was purified using an ÄKTA purifier system with a 1-ml His Trap HP column (Amersham biosciences), according to instructions of the manufacturer. The protein concentration was calculated from the absorbance at 280 nm using a computed extinction coefficient estimated to be 116,090  $M^{-1}cm^{-1}$  for the whole AIDA-I (mature AIDA-I and associated AIDA<sub>c</sub>) and 32,780  $M^{-1}cm^{-1}$  for the mature AIDA-I alone, based on the number of tryptophan and tyrosine residues (45). The purity of the purified proteins was confirmed by SDS-PAGE and staining with Coomassie blue.

**Far-UV Circular dichroism spectroscopy.** Far-UV CD spectra from 10  $\mu$ g.ml<sup>-1</sup> of the wild-type and S846L protein purified from Triton X-100 extracts were recorded in 10 mM Sodium Phosphate pH 7 in a 1-cm path length cuvette between 190 and 260 nm, using a spectropolarimeter (Jasco Spectroscopic Co. Ltd, model J-810). Multiple spectra were

obtained at 10°C interval between 30°C and 70°C using the temperature control unit. For each spectrum, three accumulations were averaged and the contribution of buffer to the measured ellipticity was subtracted. The CD spectra of 125 µg.ml<sup>-1</sup> wild-type protein purified from heat extracts were obtained similarly in a 0.1-cm path length cuvette. For the wild-type and S846L protein purified from Triton X-100 extracts, thermal denaturation was followed by monitoring the ellipticity at 218 nm of a 35 µg.ml<sup>-1</sup> solution of protein and varying the temperature between 30°C and 70°C at a speed of 5°C per min. Ellipticities were converted to Mean Residual Ellipticities (MRE).

**Auto-aggregation assay and biofilm formation.** The auto-aggregation assay was performed as described before (50). Overnight cultures from *E. coli* C600 harboring empty vector pTRC99a, pAg or pAgS846L were normalized at an OD<sub>600nm</sub> of 1.5 in 10 ml of LB in a 16x125 mm borosilicate tube. The cultures were vortexed for 10 seconds and left at 4°C for 3 h. A 200 µl sample was taken 1 cm below the surface at the beginning of the assay and after 180 min. For inhibition of auto-aggregation, a 250 µl sample of the Triton X-100 solubilized product or the heat-extracted product was added prior to the three hours incubation. Biofilm formation was monitored as described before (40), normalized cultures of *E. coli* C600 harboring empty vector pTRC99a, pAg or pAgS846L were grown without agitation for 24 h at 30°C in M9 medium containing 0,005% of proline, leucine and threonine and 100 µg.ml<sup>-1</sup> ampicillin in 96-well PVC plates (Falcon). Biofilms were stained for 15 min with crystal violet 1% and the fixed dye was solubilized by addition of ethanol:acetone (80:20). The absorbance of the dye solution was measured at 595 nm.

**Adhesion assay.** The assay was performed as previously described (2). Briefly, Hep-2 cells were grown to confluency in 24-well plates in DMEM medium without antibiotic. The cells were inoculated with 10<sup>6</sup> cfu per well of *E. coli* C600 harboring empty vector pTRC99a, pAg or pAgS849L for 3 h. Cells were washed with PBS and attached bacteria recovered with 100 µl of Triton X-100 1% in order to be plated on LB agar for numbering. As a control, wells containing only 1 ml of DMEM medium were inoculated with the same amount of bacteria. Adherence to Hep-2 cells was calculated by dividing the number of adherent bacteria with the number of bacteria found in the inoculum after 3 h of incubation.

**Invasion assay.** Hep-2 cells were infected with *E. coli* C600 harboring empty vector pTRC99a, pAg or pAgS846L for 3 h, as described above. The cells were then washed with PBS and incubated for 2 h with DMEM containing 100  $\mu\text{g}\cdot\text{ml}^{-1}$  of gentamicin and 100  $\mu\text{g}\cdot\text{ml}^{-1}$  of penicillin/streptomycin in order to kill extracellular bacteria. Cells were then lysed with 100  $\mu\text{l}$  of Triton X-100 1% in order to recover the intracellular bacteria, which were plated on LB agar for numbering. As control, bacteria were inoculated in 1 ml of medium alone and antibiotics were added. No bacteria were recovered from these wells. In other controls, cytochalasin D (0.1  $\mu\text{g}\cdot\text{ml}^{-1}$  final concentration) was added in the medium during the infection and the gentamicin treatment. The gentamicin treatment was reduced to one hour to limit the toxic effect of cytochalasin D. The presence of cytochalasin D had no effect on bacterial adhesion to Hep-2 cells (data not shown).

## RESULTS

**Outer membrane cleavage of AIDA-I is independent of its expression level.** The processing of AIDA-I after outer membrane translocation is thought to be autoproteolytic (53), but it is not known if the cleavage is intermolecular or intramolecular. To address this question we evaluated the extent of autoproteolysis under different conditions of induction, in order to assess if AIDA-I cleavage varied with protein density. Similar experiments were used to characterize the intermolecular cleavage of Hap (14, 15). We used plasmid pAg, which contains the *aidA* and *aah* genes in their original operon organization, but under the control of an IPTG-inducible promoter. We induced the expression of AIDA-I at the beginning of the exponential phase of growth and samples were collected every 10 min for the following 2 h to prepare whole cell extracts. As shown in Fig. 2A, in the earlier time points, only the mature cleaved AIDA-I is visible, not the uncleaved proprotein. This suggests that cleavage occurs even when a small amount of the adhesin is present at the cell surface. We performed a second experiment where bacteria were induced at the beginning of the exponential phase of growth with various concentrations of IPTG before obtaining normalized whole cell extracts. As shown in Fig. 2B, the predominant form of AIDA-I in all level of induction, including without induction, is the mature AIDA-I, not the uncleaved proprotein. Together, these results suggest that cleavage occurs independently from the amount of protein at the bacterial surface. This observation is consistent with an intramolecular mechanism of autoproteolysis. It should be noted, however, that this result would also be consistent with a cleavage mediated by an unknown surface protease.

**Release of the cleaved mature AIDA-I.** Previous studies have shown that the cleaved mature AIDA-I remains associated with the outer membrane by non-covalent interactions and can be released by treatment at 60°C (4). We tried to release the adhesin by other techniques known to extract proteins peripherally associated with membranes, including treatment with SDS 0.1%, Triton X-100 1%, Tween-20 0.01%, Tris-glycine pH 2, Carbonate pH 11 and treatment with urea 0.1 M or 2 M. Except for Triton X-100 1%, all these treatments failed to release the mature adhesin (data not shown). We also performed heat extraction at various temperatures between 30°C and 60°C, and we observed that only a small amount of adhesin could be released at temperatures below 42°C, while this amount increased dramatically above 50°C (data not shown). Taken together, these results suggest that the cleaved mature AIDA-I strongly interacts with the outer membrane and we could not witness its release in a wide variety of conditions.

**Co-purification of mature AIDA-I with AIDAc.** As described above, the cleaved mature AIDA-I can be released by heat treatment or with Triton X-100 1%. We wanted to purify the protein released under both conditions to compare their conformations. We introduced a six histidines and glycine (HisG) tag at the N-terminus of the mature form of AIDA-I and used immobilized metal affinity chromatography (IMAC) for purification. The result of the purification starting from heat extracts is shown in Fig. 3A and the purification from Triton X-100 extracts is shown in Fig. 3B. While the product of the purification from the heat extract was a single polypeptide, we were surprised to observe three major protein bands in the purification from Triton X-100 extracts, one with a molecular weight of 100 kDa that correspond to the mature AIDA-I and another with a molecular weight of 140 kDa that could correspond to the uncleaved proprotein. The identity of both proteins was confirmed by an immunoblot with antiserum against the cleaved mature AIDA-I (data not shown). The third protein could not be detected with these antibodies or with a serum raised against the HisG tag (data not shown). Its molecular weight of approximately 45 kDa was consistent with the size of AIDAc. We assessed the modification of mobility, under denaturing electrophoresis, when the samples were left at room temperature or incubated at 100°C. As shown in Fig. 3C, a clear difference in electrophoretic mobility for the third protein can be observed between the two samples. This property is characteristic of outer membrane proteins folded in a b-barrel conformation, including AIDAc (18, 34), thus

confirming the identity of the third protein. Overall, these results suggest that the Triton X-100 extraction co-solubilized the mature AIDA-I and AIDAc and that the two polypeptides were co-purified during IMAC. This is consistent with the strong interaction that we observed between the cleaved mature AIDA-I and the outer membrane.

To assess the conformation of the purified protein, we subjected it to far UV CD spectroscopy (27). As shown in Fig. 4A, the protein purified from the Triton X-100 extraction has a secondary structure rich in  $\beta$  sheets, as suggested by the minimum of molar ellipticity observed at 218 nm. This richness in  $\beta$  sheets is consistent with the predicted and observed b-barrel structures of outer membrane proteins, including the solved structure of the translocation domain of the NalP autotransporter (41). It is also consistent with the predicted and observed  $\beta$ -helical structures of extracellular domains of autotransporters (13, 26, 42). This result therefore strongly suggests that the purified protein is correctly folded. Furthermore, the Triton X-100 solubilized extracts could inhibit AIDA-I-mediated bacterial auto-aggregation (data not shown). Surprisingly, the protein purified in similar conditions by IMAC, but from a heat extract, is mostly devoid of secondary structure, as evidenced by the negative ellipticity at 200 nm (Fig. 4B). Consistent with this result, the heat extracts had very weak inhibitory activity (data not shown). CD spectra of the proteins were recorded at 10°C intervals between 30°C and 70°C. As shown in Fig. 4A, a dramatic shift in spectrum occurred with the Triton X-100-extracted purified protein when the temperature reached 50°C, with the resulting spectrum losing a significant portion of its structure. Since integral outer membrane proteins are notoriously heat-stable, including AIDAc (33, 34), this melting is most likely due to the denaturation of the cleaved mature AIDA-I. After heating at 60°C for a few minutes, immediately cooling the sample down to 30°C did not allow its renaturation, as judged by CD, suggesting that thermal denaturation of the mature AIDA-I is not reversible (data not shown). By contrast, the CD spectrum of the heat-extracted purified protein did not vary between 30°C and 70°C (Fig. 4B), and even at 90°C (data not shown), strengthening the conclusion that this protein was already denatured. Taken together, these results suggest that heat extraction of the cleaved mature AIDA-I causes its denaturation, whereas treatment with Triton X-100 preserves its conformation.

**Modification of amino acid at the outer membrane cleavage site.** In order to characterize the role of AIDA-I outer membrane cleavage, we constructed mutants of the cleavage site by site-directed mutagenesis. A previous study had demonstrated that this C-



terminal processing occurs between residues serine 846 and alanine 847 (53). We modified the serine 846 by a cysteine (S846C), a glycine (S846G) or a leucine (S846L) or the alanine 847 by an isoleucine (A847I), a threonine (A847T) or an arginine (A847R). We then assessed the presence of the mature form of AIDA-I in membrane fraction by western blotting using a serum raised against the cleaved mature domain (Fig. 5). The S846C, A847I and A847R mutants could not be observed. This suggests that the protein was most likely unstable and rapidly degraded. It should be noted that A847I was previously described as an unprocessed mutant and could be observed, although poorly expressed, in an *E. coli* strain deficient in outer membrane proteases (36). The sensitivity to point mutations of the cleavage region is surprising, as one could have assumed that it resides in a flexible loop. On the contrary, our results suggest that the cleavage site resides in a structured part of the protein.

The cleavage was altered to various extents in the other mutants. It was partially inhibited for S846G and A847T and almost abolished for S846L. We examined the proteins that could be heat-extracted, using SDS-PAGE and Coomassie-blue staining. As explained above, the heat treatment at 60°C only release the non-covalently associated cleaved mature AIDA-I, whereas the uncleaved proprotein remains associated with the outer membrane and cannot be extracted. As shown in Fig. 6, a small amount of cleaved mature AIDA-I from S846L could be heat-extracted, but it was a considerably lesser amount compared to wild type AIDA-I. The heat extract of S846L showed two distinct major bands, one with the normal apparent molecular weight of 100 kDa and a second with a slightly lower size. The latter disappeared in heat extracts obtained from *E. coli* strain UT5600, which lacks the OmpT and OmpP outer membrane proteases. This result suggests that when the normal autoproteolytic cleavage is inhibited, the adhesin can be cleaved at an alternative site by other outer membrane proteases such as OmpT and OmpP. This observation is consistent with previous reports that a cleavage site for OmpT exist in the surface exposed portion of AIDA<sub>c</sub> (53). This site, however, was observed only when cleaved mature AIDA-I had been removed. Our results demonstrate that this cleavage site is also accessible even in the presence of the adhesin when the autocatalytic cleavage is inhibited.

**The stability of the protein is not influenced by the lack of outer membrane cleavage.**

To assess the conformation and the stability of the uncleaved S846L mutant, we purified it and tested its conformation and thermal denaturation, monitored by CD spectroscopy as

above. As shown in Fig. 4C, the purified protein is correctly folded and its far UV CD spectrum is almost identical to that of wild-type protein. We followed more carefully the thermal denaturation of S846L and wild-type AIDA-I by monitoring the ellipticities of the proteins at 218 nm continuously while changing the temperature (Fig. 4D). The resulting curves are almost indistinguishable and yield melting temperatures of  $50.3 \pm 0.1$  °C and  $51.6 \pm 0.2$  °C for the wild-type protein and S846L, respectively. These results strongly suggest that the S846L mutant has the same stability and structure as the wild-type protein (27). A limited proteolysis with proteinase K on whole cell expressing S846L or wild type AIDA-I also indicated that the stability of both proteins is very similar (data not show).

**Lack of outer membrane cleavage does not influence the functionality of AIDA-I.** We used the S846L construct to test the role of AIDA-I outer membrane cleavage in its functionality. We first assessed if this cleavage is important for the adhesion mediated by AIDA-I. We compared the adhesive properties of the mutant with the wild-type protein in an adhesion assay on Hep2 cells. As shown in Fig. 7A, S846L had a slightly lower adhesion to cells but not statistically significant. We then tested if the outer membrane processing is important for the capacity to mediate bacterial auto-aggregation. As shown in Fig. 7B, our results clearly indicate that cleavage is not essential for bacterial auto-aggregation mediated by AIDA-I. Lastly, we assessed the effect of the processing on the ability to mediate biofilm formation (Fig. 7C). Again, our results showed that cleavage did not influence the ability to mediate biofilm formation.

**AIDA-I mediates invasion of Hep2 cell independently of its outer membrane processing.** As mentioned above, AIDA-I shares many similarities with the TibA adhesin/invasin. The major differences between these two proteins were that TibA lacks C-terminal processing and has the capacity to mediate invasion. We therefore asked whether the S846L mutation could confer invasive properties to AIDA-I. To test this hypothesis, we performed gentamicin protection invasion assays (11). As seen in Fig. 7D, S846L could indeed invade Hep2 cells but, surprisingly, we also observed that wild type AIDA-I could mediate invasion, to an extent similar to that of S846L. It should be noted, however, that the invasion we observed was markedly inferior to that conferred by TibA (11). The addition of cytochalasin D, an inhibitor of actin polymerization, reduced  $93.3 \% \pm 2.5$  (n=4) of the invasion mediated by wild-type AIDA-I and  $94.3 \% \pm 1.8$  (n=4) of invasion

mediated by S846L. This inhibition confirms that bacteria are not spontaneously resistant to gentamicin and are internalized inside Hep2 cells by a mechanism involving actin rearrangement.

## DISCUSSION

Many autotransporters are cleaved after their translocation across the outer membrane, such as the serine proteases autotransporters of *Enterobacteriaceae* (9, 19), the actin polymerizing surface protein VirG (10), or the *H. influenzae* Hap adhesin (14). The mechanisms of cleavage of autotransporters and the functions that this cleavage serves are diverse. The cleavage might be autoproteolytic (14) or due to a specific outer membrane protease (10). This processing can be required to deliver a cytotoxic polypeptide in the extracellular milieu (37), to keep a protein localized to the bacterial pole (8) or to control adherence and auto-aggregation by regulating the expression level of an adhesin (15). By contrast, the outer membrane processing of the AIDA-I autotransporter is poorly understood.

AIDA-I is cleaved in bacteria lacking the proteases OmpT, OmpP or DegP, suggesting an autoproteolytic mechanism (53). Autoproteolytic processing has been studied in many autotransporters that bear typical serine protease catalytic domains in their extracellular domains (9, 14, 19). Other autotransporter, including Ag43, have a putative aspartyl protease active site in their extracellular domain (22, 43). A catalytic site in AIDA-I that could mediate such a proteolytic processing has not been identified. In this study, we observed that outer membrane cleavage occurs independently of the level of expression of AIDA-I, suggesting that AIDA-I is cleaved by an intramolecular mechanism. To our knowledge, an intramolecular mechanism of autotransporter autoproteolysis has not yet been reported, contrasting with previous studies where an autocatalytic processing was studied and found to be intermolecular (14). It should be noted that intramolecular autoproteolysis can also occur in the absence of a typical protease catalytic site as a result of conformational stress upon protein folding (31). Further work is required to characterize the mechanism of proteolysis of AIDA-I.

Proteolytic processing of a protein often achieves a specific biological function that appears when the resulting polypeptidic fragments are separated from one another. Surprisingly, in this study we could not find conditions that could separate the cleaved mature AIDA-I from AIDA<sub>c</sub> in a folded conformation. Indeed, we found the heat-extracted

cleaved mature AIDA-I to be mostly unfolded. Triton X-100 solubilization of the mature AIDA-I co-extracted AIDAc, and both polypeptides could then be co-purified and shown to be folded. We observed that the folded AIDA-I became denatured at temperature above 50°C. This explains why the heat extraction, which is performed at 60°C, results in the release of a mostly unfolded mature protein. In previous studies, the mature forms of AIDA-I or Ag43 were purified from heat extracts by gel filtration rather than the IMAC step we used (7, 29). Presumably, the former procedure can separate correctly folded from denatured protein whereas the latter does not. However, by preventing denaturation in the first place, our procedure is likely to yield a higher amount of native protein. It should be noted that, in the conditions we used, we could not solubilize AIDA-I with Triton X-100 when it is unglycosylated (M.-E. Charbonneau and M. Mourez, unpublished data), arguing that detergent *per se* is not solubilizing AIDAc or other integral outer membrane proteins, and suggesting that the solubilization relies on the presence of carbohydrates in AIDA-I.

Our results suggest a strong association between the cleaved fragments resulting from a proteolytic processing after outer membrane translocation. This was previously observed with Ag43 (7) and questions the role of this cleavage. In order to address this issue, we engineered mutations at the cleavage site. Using the S846L mutant, we could clearly show that cleavage is not essential for the functionality of the protein in *in vitro* assays, including the capacity to mediated invasion of epithelial cell, an activity that was not previously described for AIDA-I. The *in vivo* relevance of AIDA-I mediated invasion remains to be established but it should be noted that low level of entry of many pathogenic strains of *E. coli* has previously been reported, including DAEC (25). It is often hypothesized that such mechanisms might promote persistence in target tissues (47).

In summary, the results of our study show a consistent picture of the outer membrane processing of AIDA-I and suggests that it is either fortuitous and has no further function or that it serves a purpose that is not reproduced in our *in vitro* assays. A closer examination of the mechanism of action of AIDA-I using new assays is warranted to investigate the latter hypothesis.

## ACKNOWLEDGMENTS

We are greatly indebted to Dr M. Ngeleka for the generous gift of serum raised against AIDA-I and to M. Jacques, J.-M. Betton and H. Le Moual for their comments.

This work was supported by financial contributions from the Groupe de Recherche et d'Etudes sur les Maladies Infectieuses du Porc (GREMIP), the Natural Sciences and Engineering Research Council of Canada (NSERC discovery grant 262746), the Canada Research Chair program and the Canada Foundation for Innovation (CFI project 201414).

## REFERENCES

1. **Benz, I., and M. A. Schmidt.** 1992. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. *Mol. Microbiol.* **6**:1539-46.
2. **Benz, I., and M. A. Schmidt.** 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infect. Immun.* **57**:1506-11.
3. **Benz, I., and M. A. Schmidt.** 2001. Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Mol. Microbiol.* **40**:1403-13.
4. **Benz, I., and M. A. Schmidt.** 1992. Isolation and serologic characterization of AIDA-I, the adhesin mediating the diffuse adherence phenotype of the diarrhea-associated *Escherichia coli* strain 2787 (O126:H27). *Infect. Immun.* **60**:13-8.
5. **Bilge, S. S., C. R. Clausen, W. Lau, and S. L. Moseley.** 1989. Molecular characterization of a fimbrial adhesin, F1845, mediating diffuse adherence of diarrhea-associated *Escherichia coli* to HEp-2 cells. *J. Bacteriol.* **171**:4281-9.
6. **Brandon, L. D., N. Goehring, A. Janakiraman, A. W. Yan, T. Wu, J. Beckwith, and M. B. Goldberg.** 2003. IcsA, a polarly localized autotransporter with an atypical signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed. *Mol. Microbiol.* **50**:45-60.
7. **Caffrey, P., and P. Owen.** 1989. Purification and N-terminal sequence of the alpha subunit of antigen 43, a unique protein complex associated with the outer membrane of *Escherichia coli*. *J. Bacteriol.* **171**:3634-40.
8. **d'Hauteville, H., R. Dufourcq Lagelouse, F. Nato, and P. J. Sansonetti.** 1996. Lack of cleavage of IcsA in *Shigella flexneri* causes aberrant movement and allows demonstration of a cross-reactive eukaryotic protein. *Infect. Immun.* **64**:511-7.

9. **Dutta, P. R., R. Cappello, F. Navarro-Garcia, and J. P. Nataro.** 2002. Functional comparison of serine protease autotransporters of *enterobacteriaceae*. *Infect. Immun.* **70**:7105-13.
10. **Egile, C., H. d'Hauteville, C. Parsot, and P. J. Sansonetti.** 1997. SopA, the outer membrane protease responsible for polar localization of IcsA in *Shigella flexneri*. *Mol. Microbiol.* **23**:1063-73.
11. **Elsinghorst, E. A., and D. J. Kopecko.** 1992. Molecular cloning of epithelial cell invasion determinants from enterotoxigenic *Escherichia coli*. *Infect. Immun.* **60**:2409-17.
12. **Elsinghorst, E. A., and J. A. Weitz.** 1994. Epithelial cell invasion and adherence directed by the enterotoxigenic *Escherichia coli* tib locus is associated with a 104-kilodalton outer membrane protein. *Infect. Immun.* **62**:3463-71.
13. **Emsley, P., I. G. Charles, N. F. Fairweather, and N. W. Isaacs.** 1996. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature* **381**:90-2.
14. **Fink, D. L., L. D. Cope, E. J. Hansen, and J. W. Geme, 3rd.** 2001. The *Haemophilus influenzae* Hap autotransporter is a chymotrypsin clan serine protease and undergoes autoproteolysis via an intermolecular mechanism. *J. Biol. Chem.* **276**:39492-500.
15. **Fink, D. L., and J. W. St Geme, 3rd.** 2003. Chromosomal expression of the *Haemophilus influenzae* Hap autotransporter allows fine-tuned regulation of adhesive potential via inhibition of intermolecular autoproteolysis. *J. Bacteriol.* **185**:1608-15.
16. **Fukuda, I., T. Suzuki, H. Munakata, N. Hayashi, E. Katayama, M. Yoshikawa, and C. Sasakawa.** 1995. Cleavage of *Shigella* surface protein VirG occurs at a specific site, but the secretion is not essential for intracellular spreading. *J. Bacteriol.* **177**:1719-26.
17. **Girard, V., and M. Mourez.** 2006. Adhesion mediated by autotransporters of Gram-negative bacteria: Structural and functional features. *Res. Microbiol.* **157**:407-16.
18. **Heller, K. B.** 1978. Apparent molecular weights of a heat-modifiable protein from the outer membrane of *Escherichia coli* in gels with different acrylamide concentrations. *J. Bacteriol.* **134**:1181-3.
19. **Henderson, I. R., J. Czczulin, C. Eslava, F. Noriega, and J. P. Nataro.** 1999. Characterization of pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect. Immun.* **67**:5587-96.

20. **Henderson, I. R., and J. P. Nataro.** 2001. Virulence functions of autotransporter proteins. *Infect. Immun.* **69**:1231-43.
21. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. C. Fernandez, and D. Ala'Aldeen.** 2004. Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* **68**:692-744.
22. **Henderson, I. R., and P. Owen.** 1999. The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and oxyR. *J. Bacteriol.* **181**:2132-41.
23. **Hendrixson, D. R., M. L. de la Morena, C. Stathopoulos, and J. W. St Geme, 3rd.** 1997. Structural determinants of processing and secretion of the *Haemophilus influenzae* hap protein. *Mol. Microbiol.* **26**:505-18.
24. **Hendrixson, D. R., and J. W. St Geme, 3rd.** 1998. The *Haemophilus influenzae* Hap serine protease promotes adherence and microcolony formation, potentiated by a soluble host protein. *Mol. Cell* **2**:841-50.
25. **Jouve, M., M. I. Garcia, P. Courcoux, A. Labigne, P. Gounon, and C. Le Bouguenec.** 1997. Adhesion to and invasion of HeLa cells by pathogenic *Escherichia coli* carrying the afa-3 gene cluster are mediated by the AfaE and AfaD proteins, respectively. *Infect. Immun.* **65**:4082-9.
26. **Junker, M., C. C. Schuster, A. V. McDonnell, K. A. Sorg, M. C. Finn, B. Berger, and P. L. Clark.** 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**:4918-23.
27. **Kelly, S. M., T. J. Jess, and N. C. Price.** 2005. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**:119-39.
28. **Klemm, P., R. M. Vejborg, and O. Sherlock.** 2006. Self-associating autotransporters, SAATs: functional and structural similarities. *Int. J. Med. Microbiol.* **296**:187-95.
29. **Laarmann, S., and M. A. Schmidt.** 2003. The *Escherichia coli* AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor. *Microbiology* **149**:1871-82.
30. **Lindenthal, C., and E. A. Elsinghorst.** 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect. Immun.* **67**:4084-91.

31. **Macao, B., D. G. Johansson, G. C. Hansson, and T. Hard.** 2006. Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nat. Struct. Mol. Biol.* **13**:71-6.
32. **Mainil, J. G., E. Jacquemin, P. Pohl, A. Kaeckenbeeck, and I. Benz.** 2002. DNA sequences coding for the F18 fimbriae and AIDA adhesin are localised on the same plasmid in *Escherichia coli* isolates from piglets. *Vet. Microbiol.* **86**:303-11.
33. **Mogensen, J. E., J. H. Kleinschmidt, M. A. Schmidt, and D. E. Otzen.** 2005. Misfolding of a bacterial autotransporter. *Protein Sci.* **14**:2814-27.
34. **Mogensen, J. E., D. Tapadar, M. A. Schmidt, and D. E. Otzen.** 2005. Barriers to folding of the transmembrane domain of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence. *Biochemistry* **44**:4533-45.
35. **Moormann, C., I. Benz, and M. A. Schmidt.** 2002. Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infect. Immun.* **70**:2264-70.
36. **Muller, D., I. Benz, D. Tapadar, C. Buddenborg, L. Greune, and M. A. Schmidt.** 2005. Arrangement of the translocator of the autotransporter adhesin involved in diffuse adherence on the bacterial surface. *Infect. Immun.* **73**:3851-9.
37. **Navarro-Garcia, F., A. Canizalez-Roman, J. Luna, C. Sears, and J. P. Nataro.** 2001. Plasmid-encoded toxin of enteroaggregative *Escherichia coli* is internalized by epithelial cells. *Infect. Immun.* **69**:1053-60.
38. **Ngeleka, M.** 2002. Isolation of a new *Escherichia coli* pathotype associated with diarrhea in piglets. *Can. Vet. J.* **43**:623-4.
39. **Niewerth, U., A. Frey, T. Voss, C. Le Bouguenec, G. Baljer, S. Franke, and M. A. Schmidt.** 2001. The AIDA autotransporter system is associated with F18 and stx2e in *Escherichia coli* isolates from pigs diagnosed with edema disease and postweaning diarrhea. *Clin. Diagn. Lab. Immunol.* **8**:143-9.
40. **O'Toole, G. A., L. A. Pratt, P. I. Watnick, D. K. Newman, V. B. Weaver, and R. Kolter.** 1999. Genetic approaches to study of biofilms. *Methods Enzymol.* **310**:91-109.
41. **Oomen, C. J., P. Van Ulsen, P. Van Gelder, M. Feijen, J. Tommassen, and P. Gros.** 2004. Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* **23**:1257-66.
42. **Otto, B. R., R. Sijbrandi, J. Luirink, B. Oudega, J. G. Heddle, K. Mizutani, S. Y. Park, and J. R. Tame.** 2005. Crystal Structure of Hemoglobin Protease, a Heme



- Binding Autotransporter Protein from Pathogenic *Escherichia coli*. J. Biol. Chem. **280**:17339-45.
43. **Owen, P., P. Caffrey, and L. G. Josefsson.** 1987. Identification and partial characterization of a novel bipartite protein antigen associated with the outer membrane of *Escherichia coli*. J. Bacteriol. **169**:3770-7.
  44. **Owen, P., M. Meehan, H. de Loughry-Doherty, and I. Henderson.** 1996. Phase-variable outer membrane proteins in *Escherichia coli*. FEMS Immunol. Med. Microbiol. **16**:63-76.
  45. **Pace, C. N., F. Vajdos, L. Fee, G. Grimsley, and T. Gray.** 1995. How to measure and predict the molar absorption coefficient of a protein. Protein Sci. **4**:2411-23.
  46. **Peterson, J. H., R. L. Szabady, and H. D. Bernstein.** 2006. An unusual signal peptide extension inhibits the binding of bacterial presecretory proteins to the signal recognition particle, trigger factor, and the SecYEG complex. J. Biol. Chem. **281**:9038-48.
  47. **Pizarro-Cerda, J., and P. Cossart.** 2006. Bacterial adhesion and entry into host cells. Cell **124**:715-27.
  48. **Rutherford, N., M. E. Charbonneau, F. Berthiaume, J. M. Betton, and M. Mourez.** 2006. The periplasmic folding of a cysteineless autotransporter passenger domain interferes with its outer membrane translocation. J. Bacteriol. **188**:4111-6.
  49. **Sherlock, O., U. Dobrindt, J. B. Jensen, R. Munk Vejborg, and P. Klemm.** 2006. Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. J. Bacteriol. **188**:1798-807.
  50. **Sherlock, O., M. A. Schembri, A. Reisner, and P. Klemm.** 2004. Novel Roles for the AIDA Adhesin from Diarrheagenic *Escherichia coli*: Cell Aggregation and Biofilm Formation. J. Bacteriol. **186**:8058-65.
  51. **Sherlock, O., R. M. Vejborg, and P. Klemm.** 2005. The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. Infect. Immun. **73**:1954-63.
  52. **St Geme, J. W., 3rd, M. L. de la Morena, and S. Falkow.** 1994. A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. Mol. Microbiol. **14**:217-33.

53. **Suhr, M., I. Benz, and M. A. Schmidt.** 1996. Processing of the AIDA-I precursor: removal of AIDAc and evidence for the outer membrane anchoring as a beta-barrel structure. *Mol. Microbiol.* **22**:31-42.

## FIGURE LEGENDS

**Fig. 1: Structural organization of AIDA-I.** AIDA-I is synthesized as a pre-protein cleaved at its N and C termini. The three domains resulting from these events are shown. The N-terminal signal sequence (SS) of the pre-protein is cleaved after translocation across the inner membrane, which releases the proprotein in the periplasm. After translocation across the outer membrane the proprotein is proteolytically processed, which separates the polypeptide into an N-terminal domain (mature AIDA-I) exposed to the extracellular milieu, and a membrane-embedded C-terminal domain (AIDAc).

**FIG. 2: Outer membrane cleavage of AIDA-I is independent of its level of expression.**

*A.* Whole cell extract were obtained from exponential culture of C600 harboring plasmid pAg (WT) or pTRC99a, an empty vector (-). The cultures were induced with 10  $\mu$ M of IPTG and samples of identical volumes were taken each 10 min after induction for 2 h. *B.* Whole cell extract were obtained from exponential culture of harboring plasmid pAg (WT) or an empty vector (-). The cultures were either uninduced or induced with 10 or 50  $\mu$ M of IPTG for three hours and then normalized at an OD<sub>600nm</sub> of 1.2. Proteins separated by SDS-PAGE were probed by immunoblotting with antiserum against cleaved mature AIDA-I, which allows detection of the proprotein (140 kDa) and the cleaved mature AIDA-I (100 kDa).

**FIG. 3. Co-solubilization and co-purification of AIDAc and mature AIDA-I.**

*A.* Coomassie-blue-stained SDS-PAGE gels of the heat extract of C600 harboring plasmid pAgH (lane 1) and of the purification product from this heat extract using an immobilized nickel column (lane 2). *B.* Coomassie-blue-stained SDS-PAGE gels of total membranes preparations of C600 harboring plasmid pAgH solubilized with Triton X-100 1% (lane 1) and of the purification product from this detergent extract using an immobilized nickel column (lane 2). The purified fraction shows 3 protein bands that are thought to be the uncleaved proprotein (Pro.), the cleaved mature AIDA-I, and the integral outer membrane

protein, AIDA<sub>c</sub>. *C*. The mobility of the latter is different whether the samples of purified protein is boiled at 100°C or left at room temperature prior to electrophoresis.

**FIG. 4: Thermal denaturation of AIDA-I monitored by far-UV CD.** Far-UV CD spectra of pure protein obtained from solubilization with 1% Triton X-100 (*A*, *C*) or from heat-extraction at 60°C (*B*). Multiple spectra were recorded at 10°C temperature intervals between 30°C and 70°C (light grey shades correspond to lower temperatures and dark grey shades correspond to higher temperatures) with wild-type protein (*A*, *B*) or the S846L uncleaved mutant (*C*). The ellipticities at 218 nm of the wild type protein (open squares) or the S846L mutant (filled squares) were recorded continuously with the temperature varying between 30°C and 70°C at a rate of 5°C per min (*D*).

**FIG. 5. Effect of modification of the outer membrane cleavage site of AIDA-I.** Membrane fractions were prepared from overnight cultures of C600 induced with 10 μM IPTG at OD<sub>600nm</sub> of 0.8 and harboring pAg (WT) or its mutated variants, constructed as described in the text. Proteins in the membrane fractions were separated and probed as described in Fig. 2.

**FIG. 6. Processing of AIDA-I in a protease-deficient-strain.** Strains C600 (*ompT*<sup>+</sup>/*ompP*<sup>+</sup>) or UT5600 (*ompT*<sup>-</sup>/*ompP*<sup>-</sup>) harboring an empty vector (-), pAg (WT) or pAgS846L (S846L) were grown overnight after induction with 10 μM IPTG at an OD<sub>600nm</sub> of 0.8. The cultures were concentrated in 10 mM Sodium Phosphate buffer pH 7, heated at 60°C for 20 min. The heat-extracts, containing the cleaved mature AIDA-I, were recovered from bacterial pellets by centrifugation. Proteins in these heat-extracts were then separated by SDS-PAGE and the gels stained with Coomassie blue. The star represents a second cleavage site in AIDA-I resulting from cleavage by OmpT/OmpP (see text for details).

**FIG. 7. Effect of the lack of AIDA-I outer membrane cleavage on functionality.** *A*. Adhesion assay. C600 bearing an empty vector (-), plasmid pAg (WT) or plasmid pAgS846L (S846L), were inoculated onto monolayer of confluent Hep-2 cells in 24-wells plate with 10<sup>6</sup> cfu per well. After 3 h, the cells were washed with PBS and adhering bacteria recovered, plated and counted. The adhesion is calculated by dividing the cfu of adhering bacteria recovered by the cfu found in the inoculum after 3 h of incubation. *B*.

Auto-aggregation assay. Cultures of C600 harboring an empty vector (-), plasmid pAg (WT) or pAgS846L (S846L) were standardized in 10 ml to an OD<sub>600 nm</sub> of approximately 1.5 in a culture tube and left standing at 4°C. A 200 µl sample was taken 1 cm below the surface at the beginning of the assay and after 180 min and its OD<sub>600nm</sub> was measured.

*C.* Biofilm formation. Normalized cultures of C600 harboring an empty plasmid (-), plasmid pAg (WT) or plasmid pAgS846L (S846L) were grown in M9 minimal medium for 24 h at 30°C in microtiter plates. Biofilms were stained with crystal violet. The fixed dye was solubilized with a mixture of ethanol:acetone (80:20) and the absorption of the solution measured at 595 nm.

*D.* Invasion assay. After 3 h of adhesion performed as described above, fresh medium containing gentamicin, penicillin and streptomycin was added. The cells were incubated for 2 additional hours and the surviving bacteria were recovered, plated and counted. Experiments were performed at least three times in duplicate (*A, D*) or triplicate (*B, C*) and the values, representing means ± s.e.m., show typical results.

**Table I: Primers used in this study**

Primer	Sequence
A-1F	5'- AATAAGAATTCAC TTTCTTATCACCACCAGAGATAC -3'
A-1R	5'- TTAGGATCCTTAAATGATCCTGTCAGT -3'
Aah-1F	5'- TAAAGATCTAGGAGGAATTCACCATGACTTTCTTATCACC -3'
Aah-1R	5'- TTATCTAGATTATCAGAAGCTGTATTTTATCC -3'
S846L	5'- CTTAATCCTACAAAAGAGCTCGCAGGTAATACTCTTACCGTG -3'
S846G	5'- CTTAATCCTACAAAAGAAGGCGCCGGTAATACTCTTACCGTG -3'
S846C	5'- CTTAATCCTACAAAAGAATGCGCAGGTAATACTCTTACCGTG -3'
A847I	5'- CTTAATCCTACAAAAGAATCAATTGGTAATACTCTTACCGTG -3'
A847T	5'- CTTAATCCTACAAAAGAAAGTACCGGTAATACTCTTACCGTG -3'
A847R	5'- CTTAATCCTACAAAAGAATCTAGAGGTAATACTCTTACCGTG -3'
HisG	5'- GCAGTAAATATTTTCAGGCCATCATCATCATCATGGTACAGT ATCTTCAGGAGG -3'

Figure 1

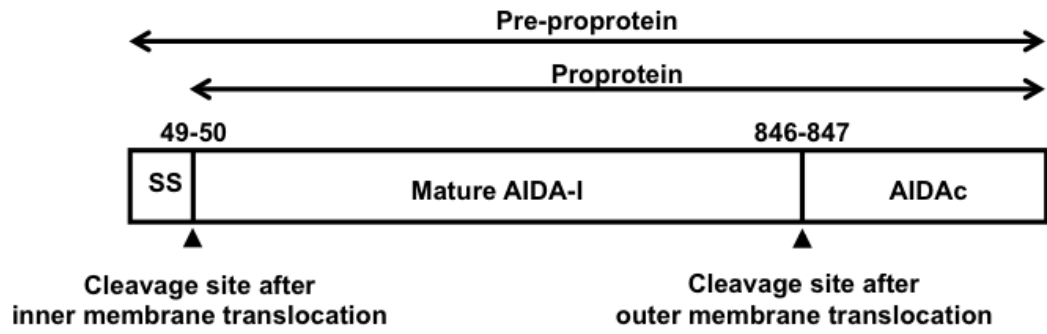


Figure 2

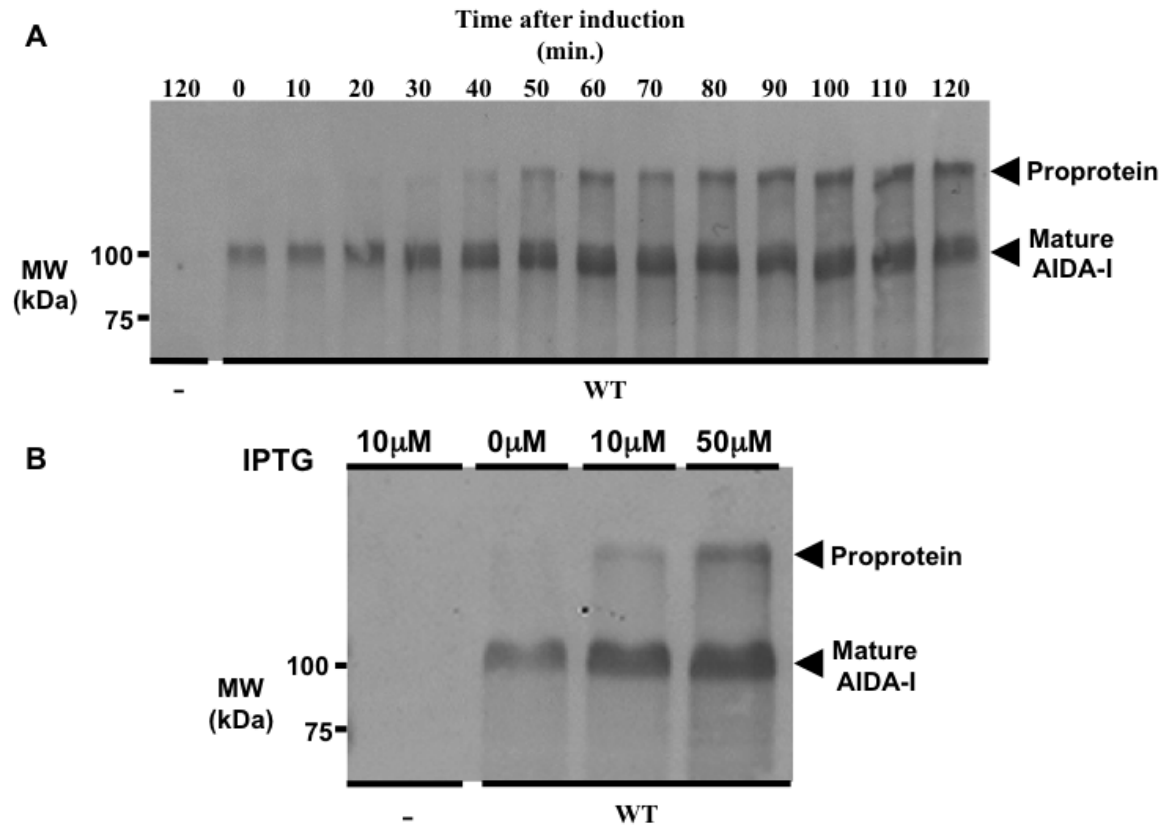


Figure 3

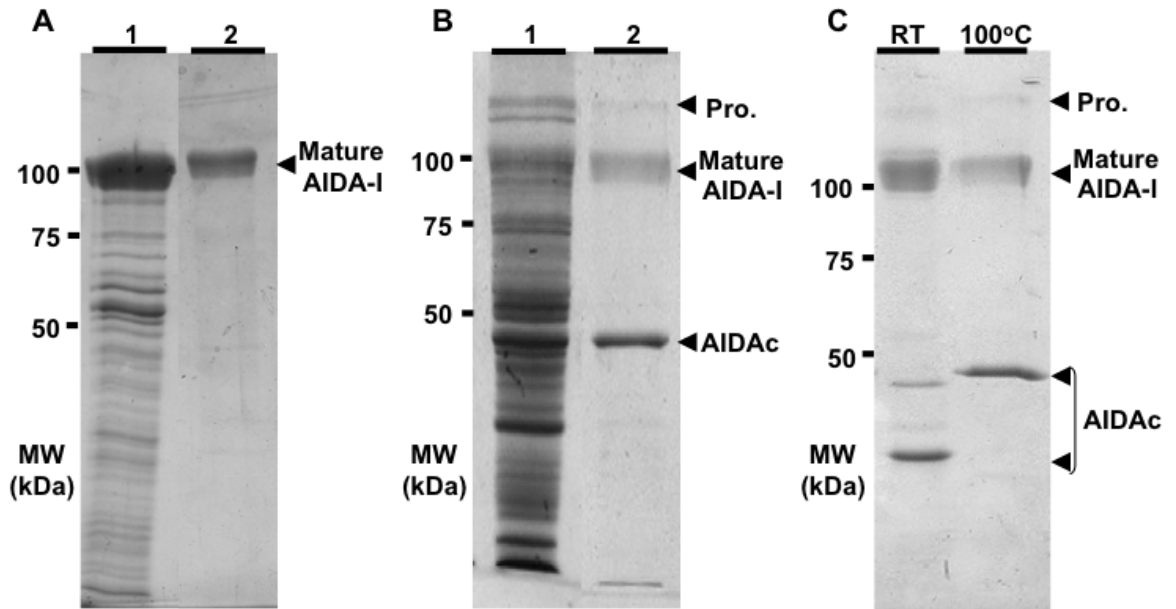




Figure 4

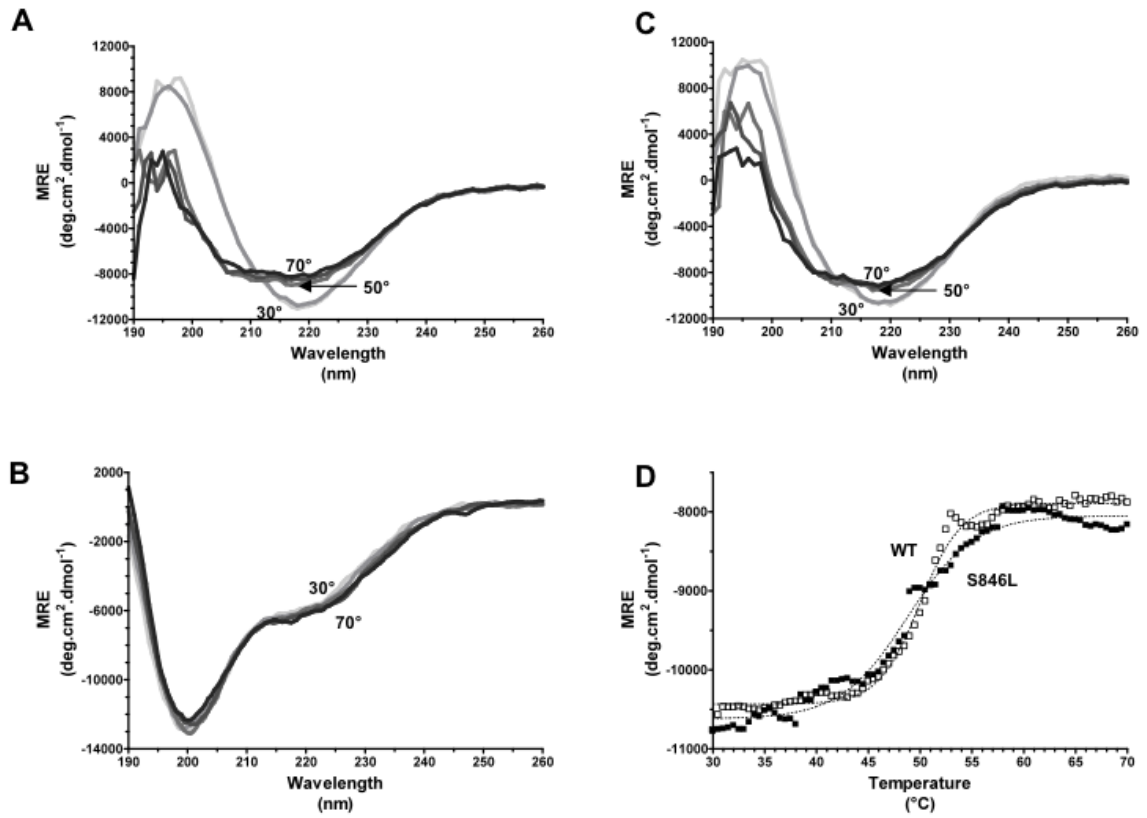


Figure 5

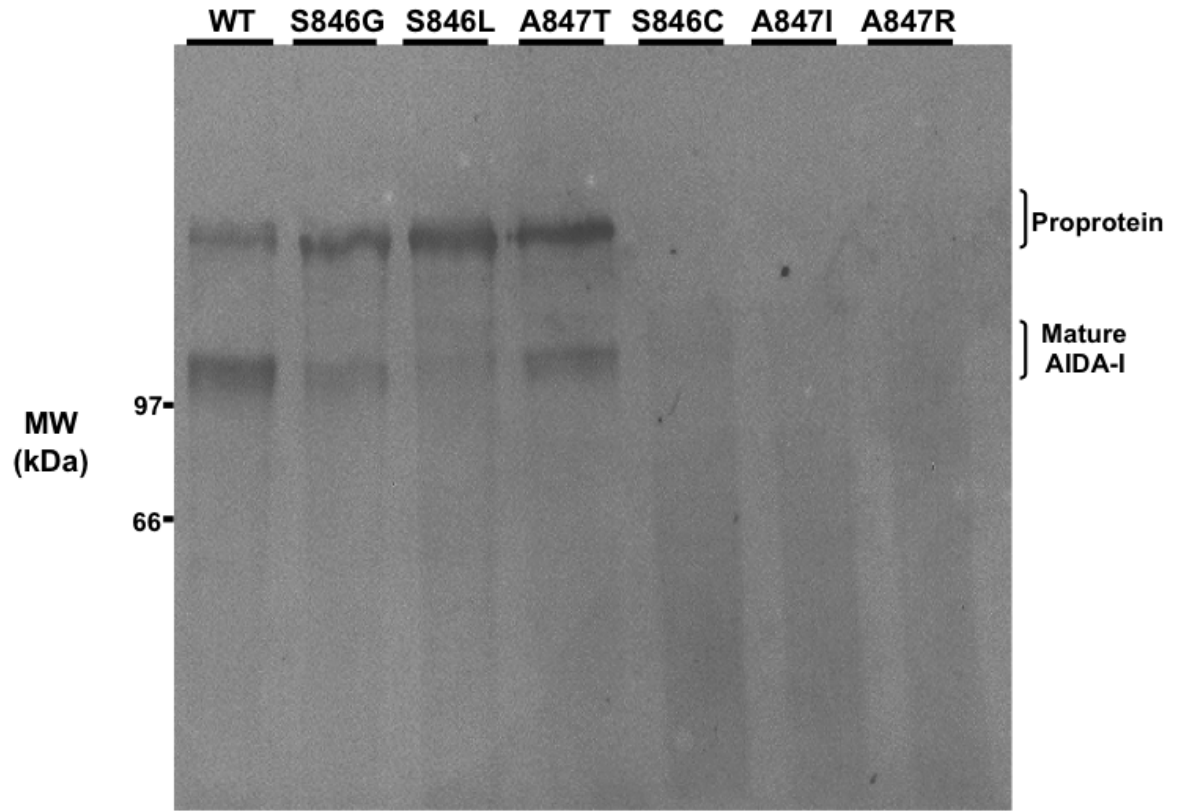


Figure 6

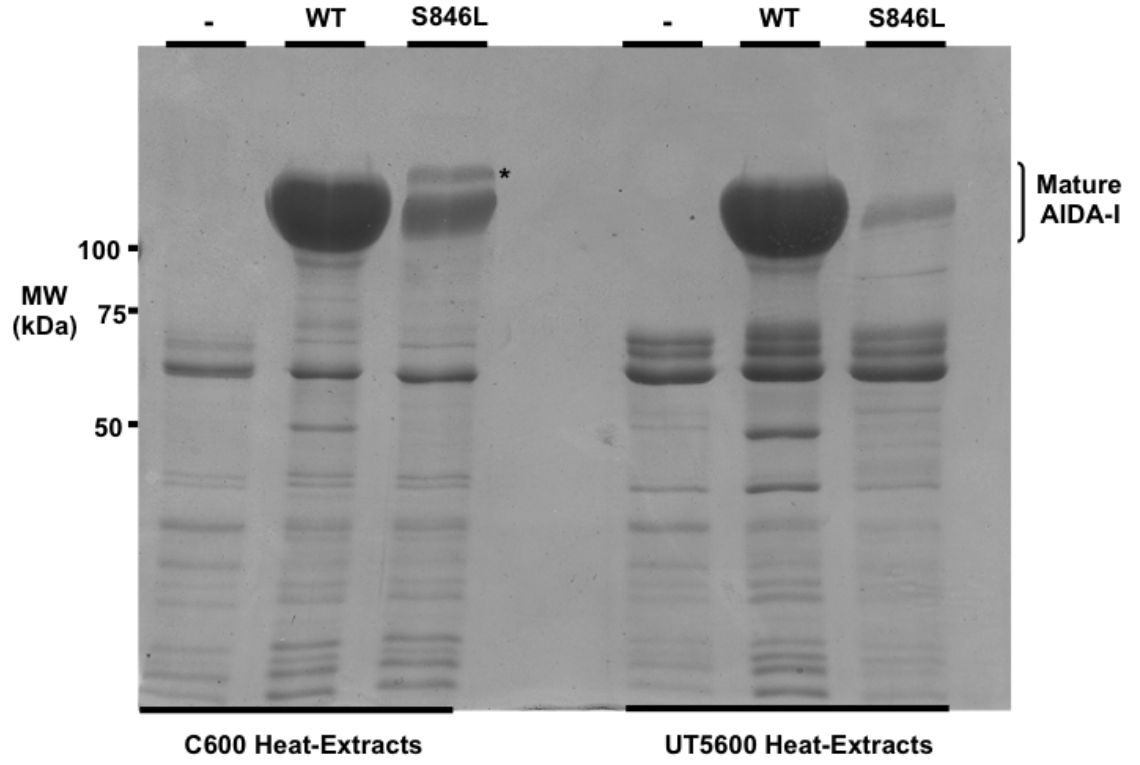
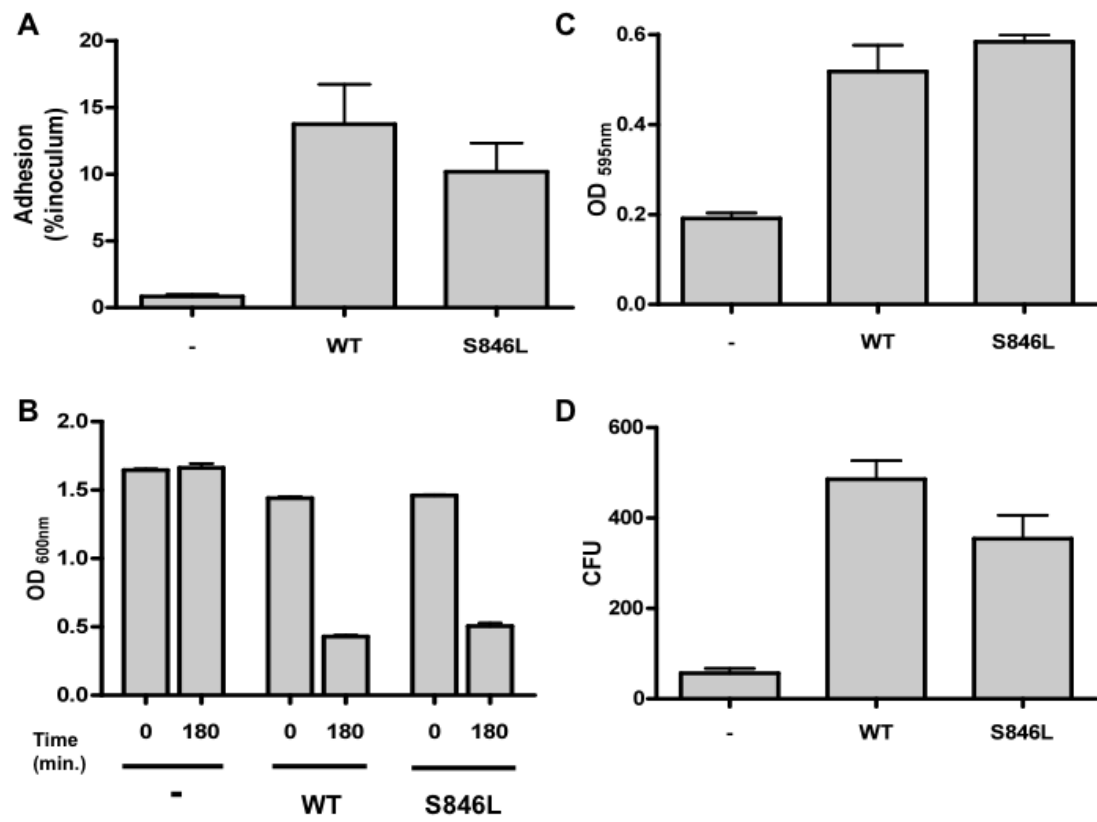


Figure 7



## ARTICLE 2

**Functional organisation of the autotransporter adhesin involved in diffuse adherence**

Journal of Bacteriology, 2007, Dec., Volume 189 (24), pages 9020-9

## Contribution de l'auteur pour cet article

En tant que premier auteur de cet article, j'ai effectué toutes les expérimentations décrites et j'ai écrit la totalité de l'article.

## Contribution scientifique de l'article

Cet article correspond à une étude portant sur la base de la multifonctionnalité de AIDA-I. Par une approche de mutagenèse aléatoire, nous avons localisé le domaine d'adhésion de AIDA-I dans la partie N-terminale de la protéine. De plus, un deuxième domaine, localisé plus centralement dans la protéine, semble être aussi impliqué dans le phénotype d'adhésion, suggérant ainsi qu'AIDA-I peut interagir avec différents récepteurs à la surface des cellules hôtes. De plus, nous avons réussi à dissocier la capacité de former des biofilms de l'agrégation bactérienne, deux phénotypes souvent considérés comme dépendants. Cette étude a donc permis d'identifier au sein du domaine passager, une région souvent vue comme une longue hélice  $\beta$  rigide, des régions impliquées dans une fonction bien précise.

**Functional organization of the autotransporter Adhesin Involved in Diffuse Adherence (AIDA-I)**

Marie-Ève Charbonneau and Michael Mourez\*

Running Title: Structure-function analysis of AIDA-I

Canada Research Chair on Bacterial Animal Diseases, Université de Montréal, Faculté de Médecine Vétérinaire, Saint-Hyacinthe, Québec, Canada

Keywords: Autotransporter, Bacterial Adhesin, AIDA-I, Outer membrane protein, *Escherichia coli*

## **Abstract**

The *Escherichia coli* Adhesin Involved in Diffuse Adherence (AIDA-I) is a multifunctional autotransporter protein that mediates bacterial aggregation, biofilm formation as well as adhesion and invasion of cultured epithelial cells. To elucidate the structure-function relationships in AIDA-I, we performed a transposon-based linker scanning mutagenesis and constructed site-directed deletions. Twenty-nine different insertions that did not affect protein expression were obtained. Eleven mutants were deficient for one or two, but not all, of the functions associated with the expression of AIDA-I. Functional characterizations of the transposon mutants and of an additional deletion mutant suggest that the N-terminal third of mature AIDA-I is involved in its binding to cultured epithelial cells. The purified product of this putative domain could bind to cultured epithelial cells, confirming the importance of this region in adhesion. We have also identified several different mutants in which invasion and adhesion are changed to different extents, and two mutants in which auto-aggregation and biofilm formation are also differently affected. These results suggest that, although conceptually linked, adhesion and invasion as well as auto-aggregation and biofilm formation are phenomena that may rely on distinct mechanisms when mediated by AIDA-I. This study brings new light on the workings of a protein belonging to an emerging family of strikingly versatile virulence factors.

## Introduction

Diarrhea-causing strains of *Escherichia coli* are responsible for many cases of gastrointestinal disease (14). Diffusely adhering *E. coli* (DAEC) is one of the six classes of *E. coli* that have been associated with diarrhea. This group is characterized by a pattern of diffuse adherence (DA) on the surface of epithelial cells (23, 29). DAEC strains have been identified from their DA on cultured epithelial cells and they appear to form a heterogeneous group. The first class of DAEC strains includes *E. coli* strains that harbor Afa/Dr adhesins, proteins that have been found to be associated with urinary tract infections and with enteric infections in infants (30). The second class of DAEC strains includes *E. coli* that express an Adhesin Involved in Diffuse Adherence (AIDA-I) (2). AIDA-I was originally identified in the strain 2787 isolated from a case of infantile diarrhea. AIDA-I has also been found in *E. coli* strains causing edema disease and post-weaning diarrhea in piglets, two major causes of economic losses in farms worldwide (21, 24, 25).

AIDA-I belongs to the family of monomeric autotransporter proteins, a branch of the type V secretion pathway (11). Most autotransporter proteins identified so far are proven or predicted virulence factors (10). AIDA-I is synthesized as a pre-proprotein of approximately 145 kDa and has a modular organization, like all autotransporters (1). The N terminus of the pre-proprotein corresponds to a *sec*-dependent sequence signal of 49 amino acids. Cleavage of the sequence signal results in the release of the proprotein in the periplasm. The proprotein comprises two different domains (34): the 100 kDa surface-exposed mature AIDA-I and the 45 kDa membrane-embedded domain, AIDAc. The latter is believed to form a pore in the outer membrane which may serve as the translocation conduit for mature AIDA-I towards the cell surface. Mature AIDA-I is cleaved after secretion, presumably by an autocatalytic event (34), but remains strongly associated with AIDAc (1, 4).

Recently, AIDA-I has been proposed to be a member of a new group of autotransporter proteins called self-associating autotransporters or SAATs (16). The group also includes Ag43, a surface protein widely distributed among pathogenic and non-pathogenic strains of *E. coli* and responsible for bacterial aggregation (27), and TibA, an adhesin/invasin of enterotoxigenic *E. coli* (5). These proteins share sequence similarities, are all glycosylated by specific heptosyltransferases that can be functionally exchanged and possess the same multiple properties likely to be important for virulence of pathogenic strains: the ability to mediate bacterial aggregation, biofilm formation, invasion and



adhesion to cultured epithelial cells (3, 9, 20, 22, 31-33). A structure-function study has been performed on Ag43 (15) and revealed that the domain responsible for the aggregation phenotype resides within the N-terminal 160 residues and is distinct from the domain allowing biofilm formation. Indeed, non-aggregative variants of Ag43 could still mediate attachment to abiotic surfaces. Ag43 can also mediate adhesion and invasion of epithelial cells (7, 31), but no region related to these phenotypes have been identified so far. In addition, this kind of analysis has never been conducted with AIDA-I or TibA. Therefore, more information is required to better understand the molecular basis of SAATs multifunctionality.

In the present report, we carried out a mutagenesis study on AIDA-I. We have used random transposon scanning mutagenesis and the construction of a domain deletion in order to identify regions in mature AIDA-I involved in the various phenotypes associated with the expression of this protein. Our data reveal that the N-terminal third of the protein is specifically involved in the attachment to cultured epithelial cells. We have also identified other mutants affected in each of the other functions of AIDA-I. The existence of mutants altered in invasion but not adhesion, and vice versa, as well as a mutant altered in auto-aggregation but not biofilm formation, and vice versa, suggests that these phenomena involve distinct mechanisms. Taken together, our results provide new insights about the mechanism of action of the SAATs group of autotransporters.

## **Materials and Methods**

**Bacterial strains and plasmids.** The *Escherichia coli* K12 strains C600 (New England Biolabs, F<sup>-</sup>, *thr1*, *leuB6*, *thi1*, *lacY1*, *supE44*, *rfbD1*, *fhuA21*) and BL21 (Statagene, F<sup>-</sup>, *ompT*, *hsdS*(r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>), *gal*) were used in this study. Plasmids pAngH and pAah allowing the expression, respectively, of AIDA-I and its specific glycosyltransferase Aah, are described elsewhere (M.-E. Charbonneau *et al.*, submitted manuscript). Plasmid pAngH is derived from the pTRC99A vector (Pharmacia Biotech) and is compatible with pAah, a pACYC184-derived plasmid. AIDA-I expressed by pAngH bears a protein tag consisting of six histidine residues and a glycine (HisG) at the N terminus of the proprotein.

To construct five amino acids insertion mutants, the pAngH plasmid was randomly mutagenized using the GPS-LS Linker Scanning System (New England Biolabs) according to the instructions of the manufacturer. The presence of the insertion was assessed by PCR.

Plasmids with insertions were transformed in the *E. coli* strain C600. DNA sequencing was used to determine the insertion sequence and its precise location.

The  $\Delta N$ ,  $\Delta C$ ,  $\Delta C2$  and  $\Delta NC$  deletions were introduced in pAngH by oligonucleotide-directed mutagenesis performed with the QuikChange II site-directed mutagenesis kit (Stratagene) using the primers listed in Table 1 and their corresponding complementary oligonucleotides, generating the constructions pAngH $\Delta N$ , (a deletion of amino acids 54 to 225) pAngH $\Delta C$  (a deletion of amino acids 224 to 637), pAngH $\Delta C2$  (a deletion of amino acids 224 to 667) and pAngH $\Delta NC$  (a deletion of amino acids 54 to 667). To generate the fusion of Glutathione S-transferase (GST) to AIDAN, a fragment of the *aidA* gene corresponding to amino acids 50 to 224 was amplified by PCR with primers that introduced *BamHI* and *XhoI* restriction sites (Table 1). The DNA fragment was digested with *BamHI* and *XhoI* and cloned at the same sites into the pGex-4T-1 vector (Amersham biosciences) to create pGex-AIDAN.

**Bacterial and cell culture growth conditions.** Bacteria containing the different plasmids were grown on Luria-Bertani (LB) agar plate or in liquid LB medium containing 100 $\mu\text{g}\cdot\text{ml}^{-1}$  ampicillin and 50  $\mu\text{g}\cdot\text{ml}^{-1}$  chloramphenicol, unless indicated otherwise. Bacterial cultures were grown at 30°C. Growth was followed by measuring turbidity as the optical density at 600 nm ( $\text{OD}_{600\text{nm}}$ ). At an  $\text{OD}_{600\text{nm}}$  of 0.8, the cultures were induced with 10  $\mu\text{M}$  isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG), unless indicated otherwise. This low concentration of IPTG was used to limit the toxicity associated with overexpression of AIDA-I. Hep-2 cells (ATCC CCL-23) were grown at 37°C with 5%  $\text{CO}_2$  in Dulbecco's modified Eagle medium (DMEM, Gibco) containing 10 mM sodium pyruvate (Sigma), 10% bovine growth serum (Hyclone), 2.5  $\mu\text{g}\cdot\text{ml}^{-1}$  fungizone and 100  $\mu\text{g}\cdot\text{ml}^{-1}$  penicillin/streptomycin (Gibco).

**Heat extraction.** Heat extracts were obtained as previously described (4). Briefly, overnight cultures from *E. coli* C600 bearing appropriate plasmids were normalized at the same  $\text{OD}_{600\text{nm}}$  in 10 ml of LB broth. Bacteria were harvested and resuspended in 150  $\mu\text{l}$  of 10 mM sodium phosphate buffer, pH 7. In order to release the mature AIDA-I, the samples were heated at 60°C for 20 minutes. The treated samples were centrifuged for 5 minutes at 12,000 x g in microfuge tubes. A 10  $\mu\text{l}$  sample from the supernatant was submitted to SDS-PAGE and stained with Coomassie blue, as described below.

**Whole cell lysates.** Overnight cultures (10 ml) were grown, normalized and centrifuged for 10 minutes at 12,000 x g in microfuge tubes and the pellets were resuspended in 100 µl of Phosphate buffered saline (PBS). All samples were processed by SDS-PAGE and immunoblotting, as described below.

**SDS-PAGE, immunoblotting and detection of glycans.** Protein samples were diluted in twice-concentrated SDS-PAGE loading buffer containing β-mercaptoethanol, and denatured by heating at 100°C for 10 minutes. The samples were then separated by SDS-PAGE on 10% or 12% acrylamide gels. The gels were either stained with Coomassie blue or transferred to polyvinylidene fluoride membranes (Millipore). Immunodetection was performed with a serum raised against the heat-extracted mature AIDA-I (a generous gift of Dr M. Ngeleka, University of Saskatchewan), diluted 1:10,000 in blocking buffer (5% skim milk, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Triton X-100). A goat anti-rabbit horseradish peroxidase (HRP) conjugated antibody (Sigma) was used as a secondary antibody, according to the instructions of the manufacturer. Alternatively, an anti-HisG HRP-coupled antibody (Invitrogen) diluted 1:5,000 in blocking buffer was used for the detection of proteins containing the HisG tag. Immune complexes were revealed using a 3,3',5,5'-Tetramethylbenzidine solution for membranes (Sigma). The detection of glycosylation was accomplished using the digoxigenin-hydrazine labeling glycan detection kit (DIG, Roche), according to the instructions of the manufacturer.

**Protein purifications.** For purification of GST fusion proteins, one liter of BL21 harboring plasmids pGex-4T-1, pGex-AIDAN or pGex-AIDAN and pAah were grown until an OD<sub>600nm</sub> of 0.4 and induced with 10 µM of IPTG for three hours. Bacteria were harvested and resuspended in 40 ml of Tris-buffered saline (50 mM Tris-HCl pH 8, 150 mM NaCl) containing lysozyme (0.4 mg.ml<sup>-1</sup> final) and EDTA pH 8 (10 mM final) and lysed with French press and an ultrasonic processor. The soluble fraction was isolated by a 30 minutes centrifugation at 16,000 x g.

For AIDA-I purification, one liter of C600 harboring plasmids pAngH and pAah were grown until an OD<sub>600nm</sub> of 0.8 and induced overnight with 10 µM of IPTG. The cells were lysed as described above and the lysate was centrifuged for 1 h in an ultracentrifuge at 250,000 x g. The membranes were resuspended in TBS containing 1% Triton X-100,

incubated for 1 hour, and centrifuged again. The solubilized membranes contained the histidine-tagged native AIDA-I.

The proteins were purified using an ÄKTA purifier system with a 5-ml glutathione sepharose column or with a 1-ml His Trap HP column (Amersham biosciences), according to the instructions of the manufacturer. The purity of the purified proteins was confirmed by SDS-PAGE and staining with Coomassie blue.

**Functional assays.** The auto-aggregation, biofilm formation, adhesion and invasion assays were performed as previously described (4). In the auto-aggregation assay cultures of *E. coli* C600 harboring the pTRC99A vector, plasmids pAah and pAngH, or plasmids pAah and one of the plasmids bearing a mutation in pAngH, were normalized in 10 ml of LB to an OD<sub>600nm</sub> of approximately 1.5 in a culture tube and left standing at 4°C. A 200 µl sample was taken 1 cm below the surface at the beginning of the assay and after 180 minutes and their OD<sub>600nm</sub> was measured. In the biofilm formation assay, normalized cultures of C600 were grown in M9 minimal medium for 24 h at 30°C in plastic microtiter plates. Biofilms were stained with crystal violet. The fixed dye was solubilized with a mixture of ethanol:acetone (80:20) and the absorption of the solution measured at 595 nm. In the adhesion and invasion assays, cultures of C600 were inoculated onto monolayers of confluent Hep-2 cells in 24-wells plate (approximately  $2.5 \times 10^5$  cells) with  $10^6$  colony-forming units per well. After 3 hours, the cells were washed with PBS and the adhering bacteria were recovered, plated and counted. For invasion, fresh medium containing gentamicin, penicillin and streptomycin was added after the three hours and an incubation of two additional hours was performed before plating. From a total number of  $1 \times 10^6$  cfu of bacteria expressing wild-type AIDA-I, the maximum of adhesion corresponds to  $7.9 \times 10^4$  cfu and the maximum of invasion corresponds to  $2.3 \times 10^2$  cfu.

All functional assays were performed at least three times in duplicate or triplicate. For each assay, the means of the results obtained with each mutant were compared to the mean obtained with the wild-type control by performing an ANOVA analysis and Dunnet post-tests using the Prism 4.0 software (Graphpad software).

**Cell-ELISA.** Hep-2 cells were grown to confluence in 96-wells plate, and fixed 15 minutes with PBS containing 2.5% paraformaldehyde and 0.2% glutaraldehyde. After one wash with PBS, the plate was blocked with 3% bovine serum albumin in PBS for two hours at

37°C. The native glycosylated AIDA-I, GST, unglycosylated GST-AIDAN or glycosylated GST-AIDAN fusion proteins were added to the cells at concentrations ranging between 300 pM and 1 µM and incubated overnight at 4°C. After extensive washes with PBS, bound proteins were detected with an HRP-coupled antibody raised against GST or against the native AIDA-I, diluted 1:10,000 in PBS (Amersham biosciences). Immune complexes were revealed using a 3,3',5,5'-Tetramethylbenzidine solution for ELISA (Sigma). Background was subtracted and the absorption intensities were normalized by dividing the absorption of each well by the maximal intensity measured on the plate. Experiments were conducted in duplicate at least twice. Binding curves and dissociation constants were obtained by nonlinear regression fitting to a one binding site hyperbola curve, using Prism 4.0 (Graphpad software).

**Immunofluorescence microscopy.** Hep-2 cells were grown to confluence in a Lab-Tek II chamber slide (Nalge Nunc International). After one wash with PBS, proteins (GST, unglycosylated GST-AIDAN or glycosylated GST-AIDAN) were added to the cells at a concentration of 1 µM and incubated at 37°C for one hour. After 4 washes with PBS, the cells were then fixed for 15 minutes with PBS containing 2.5% paraformaldehyde and 0.2% glutaraldehyde and the plate was blocked with 3% bovine serum albumin in PBS for one hour at 37°C. Bound proteins were detected with an antibody raised against GST diluted 1:1,000 in PBS-BSA 2%. Immune complexes were revealed by incubation with a donkey anti-goat IgG rhodamine red conjugate and examining the microscopic slides with a fluorescence and phase contrast Leica DMI4000B microscope (Meyer instruments).

## Results

**Generation of five amino acids insertions in AIDA-I.** For the analysis of mature AIDA-I, we used a Tn7-derived transposon system that results in five amino acids insertions in the protein. The mutagenesis was executed on the plasmid pAngH, which allows the expression of AIDA-I under the control of an IPTG-inducible promoter. In addition, AIDA-I expressed by this construct bears a HisG tag located at the N-terminal end of the proprotein. The tag did not affect the expression of AIDA-I and bacteria expressing the tagged protein had identical adhesion, invasion, biofilm formation and auto-aggregation activities, compared with bacteria expressing the wild-type untagged protein (data not shown). Glycosylation of the protein is important to ensure the normal conformation of the protein and, as a

consequence, for the cell-binding properties of AIDA-I (3, and M-E. Charbonneau *et al.* submitted for publication). We therefore used the plasmid pAah, which allowed the expression of Aah, the glycosyltransferase modifying AIDA-I (3). The two plasmids, pAngH and pAah carry compatible origins of replication and were introduced in the *E. coli* strain C600.

The mutagenesis yielded 60 different insertions distributed throughout mature AIDA-I. Among these insertions, 29 did not affect protein expression, as assessed by an immunoblotting on whole cell lysates with an antibody against the HisG tag (data not shown). The insertion sites and the inserted sequences in these mutants are given in Supplementary Table 1. The remaining thirty-one insertions that disrupted protein expression were discarded from our study. One reason for the latter is that the mutagenesis procedure leaves a *PmeI* restriction site, the translation of which will result in a TAA stop codon in 2 of the 6 possible frames. Many of the insertions that result in an absence of protein expression are probably due to this introduction of a stop codon and would therefore not be informative. Alternatively, some of the insertion mutants could result in an unstable protein.

As detailed below, 11 of the 29 insertion mutants are affected in their function. We examined in greater details the expression of these mutants, along with a mutant that is unaffected by the insertion (I8) as a control. As shown in Fig. 1, the 100-kDa mature AIDA-I can be observed in whole cell lysates of all mutants and it is correctly located at the cell surface, as shown by the release of mature AIDA-I in the supernatant after a brief heating at 60°C (1). Moreover, mature AIDA-I for all mutants appears at a molecular weight that represents the glycosylated form of the protein and it can be detected with the anti-AIDA-I antiserum, which only recognizes the glycosylated protein (3). Based on those observations we can conclude that all insertion mutants express glycosylated AIDA-I, although it is still possible that the extent of glycosylation might slightly differ between mutants. This is in agreement with our recent finding that glycosylation is highly heterogeneous and occurs at many sites throughout AIDA-I (M-E. Charbonneau, *et al.* submitted for publication). These results therefore suggest that the insertions did not affect the biogenesis, including the glycosylation status, or the structure of AIDA-I, and, in other words, that these sites are structurally permissive for the insertions. For the I2 and I20 mutants, however, two distinct additional bands in the whole cell lysates reacted with the antibody against mature AIDA-I or against the HisG tag (Fig. 1A and B). These two

polypeptides could also be released in the culture supernatant after brief heat treatment (Fig. 1C). N-terminal sequencing was used to identify the cleavage sites. The cleavage for both mutants occurs at the transposon insertion site, between a lysine and a histidine (Supplementary Table 1). Since these additional cleavage sites were introduced by the inserted sequences and the proteins only partly processed, we considered that the structural integrity of these proteins was most likely not dramatically affected.

**Effect of the insertions on AIDA-I.** To determine the effects of the insertions on the functionality of the protein, the activity of each mutant was tested using four distinct functional assays: auto-aggregation, biofilm formation, adhesion and invasion assays. Fig. 2 shows the results for all the mutants with a deficiency in one or more functions. The mutants I8, which is as functional as the wild-type protein, was included as a control in the results shown in Fig. 2. The other mutants that are not represented are as functional in all assays as the wild-type protein.

Eleven of the 29 insertions caused a defect in one or more functions associated with the expression of AIDA-I. The positions of the insertions, as well as the results of the functional characterizations, are summarized in Fig. 3. Mutations in I1, I2, I4, I5 and I6, all located in the N-terminal third of the mature AIDA-I, caused a defect in adhesion. As a logical consequence, most of these mutations caused a defect in the ability to invade cultured epithelial cells, with the exception of mutant I4 that was not statistically different from the wild-type (Fig. 2C). These results suggest that the N-terminal region of mature AIDA-I is involved in adhesion. Three other insertions in the C-terminal half of the protein (mutants I20, I23 and I24) also caused a diminution in the ability to mediate adhesion. It is tempting to speculate that the region encompassing the position of these mutants could therefore represent a second cell-binding domain in AIDA-I but, alternatively, the mutation could indirectly affect the same binding site affected by the mutations in the N-terminal part of the protein. We also observed that bacteria expressing three different mutants, I9, I12 and I13, could adhere to cultured cells as well as the wild-type protein but were unable to mediate invasion. This observation therefore argues that adhesion and invasion involve different mechanisms. Consistent with this view, but unexpectedly, we observed that invasion mediated by the I4, I23 and I24 mutants was not significantly affected, whereas their ability to mediate adhesion was reduced, sometimes dramatically (see mutant I23, Fig. 2C).

Only one mutant, I24, located in the C-terminal part of mature AIDA-I, was unable to mediate bacterial aggregation, but this mutant could make biofilm as well as the wild-type protein (Fig. 2A and 2B). This result suggests that biofilm formation and bacterial aggregation are due to distinct mechanisms, as was previously observed with Ag43 (15). Consistent with this view, one mutant in the N-terminal part of mature AIDA-I, I9, caused a slight diminution in the ability to form biofilm but mediated bacterial aggregation as well as the wild-type protein (Fig. 2A and 2B).

**Deletion of a putative cell-binding domain in mature AIDA-I.** Five insertion mutants clustered in the N-terminal third of mature AIDA-I affected the adhesion mediated by the protein, suggesting the existence of a cell-binding domain in this region. In order to complement these results, we constructed a deletion of the N-terminal third of mature AIDA-I between residues G54 and A225 (the former corresponds to the fourth residue of mature AIDA-I and the latter corresponds to the transposon insertion site in mutant I7, see Fig. 4A).

The correct expression and biogenesis of the  $\Delta N$  mutant was checked by probing whole cell lysates with antibodies against the HisG tag. As shown in Fig. 4B, a specific polypeptide slightly above 75 kDa was apparent in proteins extracted from bacteria expressing the  $\Delta N$  mutant, consistent with the expected size of the mature  $\Delta N$  mutant (63 kDa), with glycosylation. This hypothesis was confirmed by the detection of sugar using a glycan detection kit. The reaction of the  $\Delta N$  mutant with the digoxigenin-hydrazine label was identical to that of the wild-type protein (Fig. 4D). In another study, we have identified 16 residues of mature AIDA-I which can be modified by heptose molecules (M-E. Charbonneau, *et al.* submitted for publication). Only four of these sites are located between G54 and A225. The latter can therefore explain the fact that the  $\Delta N$  mutant is still glycosylated. Additionally, we performed a heat-extraction of bacteria expressing the  $\Delta N$  mutant to verify the correct location of the protein at the cell surface (Fig. 4C). Again, a polypeptide of the expected size was apparent, without any degradation product. These assays show that the  $\Delta N$  mutant was as well expressed, glycosylated and stable as the wild-type protein.

To assess the effects of this deletion on the functionality of AIDA-I, the four functional assays performed above were conducted again (Fig. 5). The deletion of the N-



terminal third of mature AIDA-I did not affect the ability of the protein to mediate bacterial aggregation or biofilm formation (Fig. 5A and B). This suggests that the regions of AIDA-I involved in biofilm formation and auto-aggregation are not located in the N-terminal third of the mature protein. However, the  $\Delta N$  mutant was considerably less able to mediate binding to cultured epithelial cells (Fig. 5C). This is in agreement with the results from the linker scanning mutagenesis that indicated that mutants I1, I2, I4, I5 and I6 were affected in their ability to mediate adhesion. This appears in contradiction with the hypothesis of the existence of a secondary cell-binding site. It should be noted however that the invasion of the  $\Delta N$  mutant was marginally lower than that of wild-type AIDA-I (Fig. 5D). This result shows that a low level of cell-binding ability remains in the  $\Delta N$  mutant, consistent with the hypothesis of the existence of two regions important for adhesion. This difference in the extent by which adhesion and invasion are affected is also similar to the observation made with the I4, I23 and I24 mutants (Fig. 2C).

We constructed three additional deletions in mature AIDA-I, corresponding to deletion in the C-terminal portion of the repeats (between G224 and A637 or between G224 and A667) or of all the repeats (between Gly54 and A667). These mutants were visibly less expressed than the wild-type protein or the  $\Delta N$  mutant or not at all (data not shown), suggesting that the biogenesis, the stability at the cell surface and/or the global conformation of these mutated proteins may be disturbed. It is therefore likely that the C-terminal repeats of mature AIDA-I play a more important role in the biogenesis or the structural integrity of the protein. Since these defects were likely to influence the functional assays of the mutants, we decided not to test their functions.

**The N-terminal domain of mature AIDA-I, glycosylated or not, is sufficient to mediate binding to cultured epithelial cells.** To confirm that the N-terminal third of mature AIDA-I is involved in the adhesion to epithelial cells, we fused this region, corresponding to residues A50 to G224, to the Glutathione S-transferase (GST). We purified by affinity chromatography the GST protein alone as control and the unglycosylated and glycosylated forms of GST-AIDAN fusion protein by co-expression with or without the Aah glycosyltransferases (Fig. 6A). The GST-AIDAN fusion encompasses 4 glycosylation sites we identified. The glycosylation status of this protein was confirmed by the glycan detection kit (Fig. 6B). In order to demonstrate the binding properties of the N-terminal

third of the mature AIDA-I, we developed a cell-based ELISA. As shown in Fig. 6C, unglycosylated and glycosylated GST-AIDAN fusion proteins bound in a saturable manner to the Hep-2 cells with an equilibrium dissociation constant of 84 nM and 72 nM, respectively. By comparison, we observed that the whole glycosylated AIDA-I could bind to the cells with a dissociation constant of 3 nM. This result is in agreement with the previous dissociation constant of 2 nM obtained in another study performed with the mature protein (19). The higher dissociation constant obtained with the fusion proteins (84 and 72 nM, compared to 2 or 3 nM for whole AIDA-I) could be due to the presence of the GST moiety blocking accessibility to cellular receptors, by the difficulty to purify a fusion protein as stable as the native protein or by the possible presence of two binding sites in mature AIDA-I, as suggested by the result of our mutagenesis.

In order to qualitatively assess the adhesion properties of the N-terminal region of mature AIDA-I and to confirm the results of the cell-based ELISA, we have used an immunofluorescence microscopy technique to visualize the binding of unglycosylated or glycosylated GST-AIDAN to live cells using an antibody raised against GST. As shown in Fig. 7, unglycosylated and glycosylated GST-AIDAN can bind specifically to live cells in a similar manner, whereas GST alone gives no signal. This result highlights again the importance of the N-terminal domain of AIDA-I in adhesion to epithelial cells and suggest further that glycosylation is not important in adhesion.

## **Discussion**

In this study, we generated 29 different mutants by linker scanning mutagenesis. The mutants were distributed along the full length of mature AIDA-I. However, as shown in Fig. 8A, 27 of the 29 insertions were located in the first 700 amino acids of the protein, a region of mature AIDA-I consisting in approximately 35 imperfect repeats of a 19-amino acid sequence shared with the other SAATs (16). By contrast, only two insertions were found in the last 200 amino acids of mature AIDA-I, suggesting that there was a bias against insertions in this part of the protein. Since we discarded from our study the insertions that disrupted protein expression, it is tempting to speculate that the region outside the repeats may be important for the structural integrity of the protein.

The structures of extracellular domains of many autotransporters are thought to consist in a  $\beta$  helix. A recent study proposed structural models for various autotransporters helix coils, which consist in 3  $\beta$  strands (13). We used the model proposed for TibA (Fig.

8B), since the consensus sequence of its repeats is identical to that of AIDA-I, in order to map the transposon insertion sites in the AIDA-I mutants on the model (Fig. 8C). Interestingly, almost none of the insertion occurring in the loops connecting the strands affected function. This might be due to the fact that the loops can easily accommodate additional amino acids, compared with strands. The previous structures of extracellular domains of autotransporters show that the  $\beta$  helix are stabilized by stacked amino acid residues (6, 26). An insertion in a strand might therefore displace a part of this strand either before or after the position of the stacked residues, which would act as an “anchor”. The inserted residues would then replace part of the strand and “bulge out” of the  $\beta$  helix. The insertion could additionally have long-distance effects by destabilizing the whole  $\beta$  helix. It is likely, however, that in the latter case the stability of the whole protein would be severely affected. It should be noted that different sequences inserted at the same location, such as mutants I8 and I9 for instance, gave different phenotypes in the assays. Similarly, the I3 mutant has an insertion between amino acids G149 and H150 and behave like the wild-type strain whereas the neighboring insertion mutants I2 (L148-G149) and I4 (H150-A151) are deficient for adhesion. These observations suggest that the inserted sequence, as well as the location of the insertion, influence the functionality of the resulting mutant.

As shown in Fig. 8C, the  $\beta$ 2 strands seems to harbor more of the insertions that did not affect function, whereas the  $\beta$ 1 strands seems more important for the functionality of the protein, as five of the six insertions in these strands disrupted one or more functions related to the expression of AIDA-I. This observation raises the possibility that the functionality of AIDA-I might be localized predominantly on one of the faces of the  $\beta$  helix. In this respect, we noticed that the mutants affecting adhesion and/or invasion are clustered in strands  $\beta$ 1 and, to a lesser extent, strands  $\beta$ 2 and  $\beta$ 3, whereas the mutants affected in biofilm or auto-aggregation are on strands  $\beta$ 2. This could suggest that the interactions with eukaryotic cells involves primarily the face of the  $\beta$  helix formed by the  $\beta$ 1 strands, whereas auto-aggregation or biofilm formation requires the face formed by the  $\beta$ 2 strands.

The main conclusion of our study is that a cell-binding domain resides within the N-terminal third of mature AIDA-I. First, we identified five different insertions in this N-terminal region that result in a deficiency in adhesion to cultured epithelial cells. Second, AIDA-I with a deletion of the amino acids between G54 and A225 is unable to mediate adhesion to Hep-2 cells. Lastly, the N-terminal region (between A50 and G224) fused to GST can efficiently bind to epithelial cells as assessed by a cell-based ELISA and by

immunofluorescence microscopy. The fusion protein could bind whether it is glycosylated or not. The observation that glycosylation is not required for binding is in agreement with our recent finding that glycosylation of AIDA-I is required to ensure the normal conformation of the protein, and therefore might only indirectly affect the adhesion to epithelial cells (M-E. Charbonneau *et al.*, submitted for publication). Based upon the known structures of others autotransporters (6, 12, 26), the N-terminal location of the primary binding domain of AIDA-I could make the interaction with a host cell receptor easier, as this part of the protein is farther away from the cell surface. The location of a cell-binding domain at the N-terminal end of a bacterial adhesin has been reported before, as for instance in the case of the *Yersinia enterocolitica* trimeric autotransporter adhesin YadA (28).

We have also identified in the C-terminal half of mature AIDA-I three other insertion mutants that exhibited a reduction in their adhesion phenotype. Our results therefore suggest that the protein possesses two binding domains, one in the N-terminal third and one in the C-terminal half. The presence of different binding domains in the same adhesin has already been described for instance in trimeric autotransporters such as the *Haemophilus influenzae* Hia (18) as well as in the *E. coli* Dr adhesin (17). As mentioned above, we noticed that insertions affecting adhesion were mostly located in the predicted  $\beta 1$  strands and to a lesser extent in the predicted  $\beta 3$  strands. The two faces of the  $\beta$  helix formed by these strands might therefore harbor the two different binding sites. Alternatively, the insertions in the C-terminal half of the protein could indirectly affect the N-terminal cell-binding domain. For instance the N- and C-termini of the repeats might be spatially close in the three-dimensional structure. More work is required at this time to distinguish between these possibilities.

Our results further suggest that invasion and adhesion are relying on different mechanisms. Indeed, we have identified three insertion mutants, I9, I12 and I13, which mediated binding to epithelial cell as well as the wild-type protein but are unable to mediate invasion. Invasion mediated by AIDA-I is therefore likely to be not just the effect of a passive uptake by cultured epithelial cells, but seems rather to rely on a specific mechanism. Bacteria expressing the I4, I23 and I24 insertion mutants or the  $\Delta N$  deletion mutant have impaired adhesion but only slightly reduced invasion. These results suggest that the binding to a receptor involved in adhesion could be different from the interactions responsible for invasion. Similarly, it was recently observed that variants of the YadA

trimeric autotransporter with reduced binding to laminin had a higher binding to fibronectin and an increase in integrin-mediated invasion (8).

Lastly, we obtained 2 mutants (I9 and I24) for which the abilities to mediate auto-aggregation and biofilm formation were not correlated: a non-aggregative mutant could mediate biofilm formation and vice versa. The separation of these two a priori related phenotypes has also been noted for Ag43 (15). Collectively, these studies therefore support the idea that biofilm formation mediated by AIDA-I involves more than the ability of bacterial cells to interact with one another.

A more precise picture of the AIDA-I autotransporter is emerging from our study. Given the similarity that AIDA-I shares with Ag43 and TibA, our results are likely to apply to these and other putative SAAT proteins and shed new light on how these versatile proteins participate in the pathogenesis of diarrheagenic *E. coli*.

### **Acknowledgements**

This work was supported by financial contributions from the Groupe de Recherche et d'Etudes sur les Maladies Infectieuses du Porc (GREMIP), the Natural Sciences and Engineering Research Council of Canada (NSERC discovery grant 262746), the Canada Research Chair program and the Canada Foundation for Innovation (CFI project 201414), and a graduate fellowship to M.-E. C. from the Fonds de Recherche sur la Nature et les Technologies du Québec (FQRNT #114663).

### **References**

1. **Benz, I., and M. A. Schmidt.** 1992. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. *Mol Microbiol* **6**:1539-46.
2. **Benz, I., and M. A. Schmidt.** 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infect Immun* **57**:1506-11.
3. **Benz, I., and M. A. Schmidt.** 2001. Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Mol Microbiol* **40**:1403-13.

4. **Charbonneau, M. E., F. Berthiaume, and M. Mourez.** 2006. Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I). *J Bacteriol* **188**:8504-12.
5. **Elsinghorst, E. A., and J. A. Weitz.** 1994. Epithelial cell invasion and adherence directed by the enterotoxigenic *Escherichia coli* *tib* locus is associated with a 104-kilodalton outer membrane protein. *Infect Immun* **62**:3463-71.
6. **Emsley, P., I. G. Charles, N. F. Fairweather, and N. W. Isaacs.** 1996. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature* **381**:90-2.
7. **Fexby, S., T. Bjarnsholt, P. O. Jensen, V. Roos, N. Hoiby, M. Givskov, and P. Klemm.** 2007. Biological Trojan horse: Antigen 43 provides specific bacterial uptake and survival in human neutrophils. *Infect Immun* **75**:30-4.
8. **Heise, T., and P. Dersch.** 2006. Identification of a domain in *Yersinia* virulence factor YadA that is crucial for extracellular matrix-specific cell adhesion and uptake. *Proc Natl Acad Sci U S A* **103**:3375-80.
9. **Henderson, I. R., M. Meehan, and P. Owen.** 1997. Antigen 43, a phase-variable bipartite outer membrane protein, determines colony morphology and autoaggregation in *Escherichia coli* K-12. *FEMS Microbiol Lett* **149**:115-20.
10. **Henderson, I. R., and J. P. Nataro.** 2001. Virulence functions of autotransporter proteins. *Infect Immun* **69**:1231-43.
11. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. C. Fernandez, and D. Ala'Aldeen.** 2004. Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* **68**:692-744.
12. **Junker, M., C. C. Schuster, A. V. McDonnell, K. A. Sorg, M. C. Finn, B. Berger, and P. L. Clark.** 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**:4918-23.
13. **Kajava, A. V., and A. C. Steven.** 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**:306-15.
14. **Kaper, J. B., J. P. Nataro, and H. L. Mobley.** 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**:123-40.

15. **Klemm, P., L. Hjerrild, M. Gjermansen, and M. A. Schembri.** 2004. Structure-function analysis of the self-recognizing Antigen 43 autotransporter protein from *Escherichia coli*. *Mol Microbiol* **51**:283-96.
16. **Klemm, P., R. M. Vejborg, and O. Sherlock.** 2006. Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol* **296**:187-95.
17. **Korotkova, N., E. Cota, Y. Lebedin, S. Monpouet, J. Guignot, A. L. Servin, S. Matthews, and S. L. Moseley.** 2006. A subfamily of Dr adhesins of *Escherichia coli* bind independently to decay-accelerating factor and the N-domain of carcinoembryonic antigen. *J Biol Chem* **281**:29120-30.
18. **Laarmann, S., D. Cutter, T. Juehne, S. J. Barenkamp, and J. W. St Geme.** 2002. The *Haemophilus influenzae* Hia autotransporter harbours two adhesive pockets that reside in the passenger domain and recognize the same host cell receptor. *Mol Microbiol* **46**:731-43.
19. **Laarmann, S., and M. A. Schmidt.** 2003. The *Escherichia coli* AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor. *Microbiology* **149**:1871-82.
20. **Lindenthal, C., and E. A. Elsinghorst.** 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect Immun* **67**:4084-91.
21. **Mainil, J. G., E. Jacquemin, P. Pohl, A. Kaeckenbeeck, and I. Benz.** 2002. DNA sequences coding for the F18 fimbriae and AIDA adhesin are localised on the same plasmid in *Escherichia coli* isolates from piglets. *Vet Microbiol* **86**:303-11.
22. **Moormann, C., I. Benz, and M. A. Schmidt.** 2002. Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infect Immun* **70**:2264-70.
23. **Nataro, J. P., I. C. Scaletsky, J. B. Kaper, M. M. Levine, and L. R. Trabulsi.** 1985. Plasmid-mediated factors conferring diffuse and localized adherence of enteropathogenic *Escherichia coli*. *Infect Immun* **48**:378-83.
24. **Ngeleka, M.** 2002. Isolation of a new *Escherichia coli* pathotype associated with diarrhea in piglets. *Can Vet J* **43**:623-4.
25. **Niewerth, U., A. Frey, T. Voss, C. Le Bouguenec, G. Baljer, S. Franke, and M. A. Schmidt.** 2001. The AIDA autotransporter system is associated with F18 and stx2e in

- Escherichia coli* isolates from pigs diagnosed with edema disease and postweaning diarrhea. Clin Diagn Lab Immunol **8**:143-9.
26. **Otto, B. R., R. Sijbrandi, J. Luirink, B. Oudega, J. G. Heddle, K. Mizutani, S. Y. Park, and J. R. Tame.** 2005. Crystal Structure of Hemoglobin Protease, a Heme Binding Autotransporter Protein from Pathogenic *Escherichia coli*. J Biol Chem **280**:17339-45.
  27. **Owen, P., P. Caffrey, and L. G. Josefsson.** 1987. Identification and partial characterization of a novel bipartite protein antigen associated with the outer membrane of *Escherichia coli*. J Bacteriol **169**:3770-7.
  28. **Roggenkamp, A., K. Ruckdeschel, L. Leitritz, R. Schmitt, and J. Heesemann.** 1996. Deletion of amino acids 29 to 81 in adhesion protein YadA of *Yersinia enterocolitica* serotype O:8 results in selective abrogation of adherence to neutrophils. Infect Immun **64**:2506-14.
  29. **Scaletsky, I. C., M. L. Silva, and L. R. Trabulsi.** 1984. Distinctive patterns of adherence of enteropathogenic *Escherichia coli* to HeLa cells. Infect Immun **45**:534-6.
  30. **Servin, A. L.** 2005. Pathogenesis of Afa/Dr diffusely adhering *Escherichia coli*. Clin Microbiol Rev **18**:264-92.
  31. **Sherlock, O., U. Dobrindt, J. B. Jensen, R. Munk Vejborg, and P. Klemm.** 2006. Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. J Bacteriol **188**:1798-807.
  32. **Sherlock, O., M. A. Schembri, A. Reisner, and P. Klemm.** 2004. Novel Roles for the AIDA Adhesin from Diarrheagenic *Escherichia coli*: Cell Aggregation and Biofilm Formation. J Bacteriol **186**:8058-65.
  33. **Sherlock, O., R. M. Vejborg, and P. Klemm.** 2005. The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. Infect Immun **73**:1954-63.
  34. **Suhr, M., I. Benz, and M. A. Schmidt.** 1996. Processing of the AIDA-I precursor: removal of AIDAc and evidence for the outer membrane anchoring as a beta-barrel structure. Mol Microbiol **22**:31-42.



## Figure legends

### Fig. 1: Effect of insertions on the expression of AIDA-I

Whole cell lysates were obtained from overnight cultures of C600 harboring an empty vector (-), plasmid pAngH (WT), or one of the plasmids bearing a five amino acids insertion in pAngH. Bacteria harboring pAngH plasmids also contained the pAah plasmid allowing the production of the AIDA-I specific glycosyltransferase, Aah. Proteins separated by SDS-PAGE were probed by immunoblotting with an antiserum against mature AIDA-I (*A*) or with an antibody against the HisG tag (*B*) which allows detection of the proprotein (Pro., 132 kDa, closed circle) and the mature AIDA-I (100 kDa, closed arrowhead). *C*. Strain C600 harboring the same plasmids were heat-extracted at 60°C for 20 min. The heat-extracts were resolved by SDS-PAGE and stained with Coomassie blue. The stars represent additional cleavage sites in AIDA-I resulting from the inserted sequence (see text for details).

### Fig. 2: Effect of insertion mutants on the functionality of AIDA-I

Bacteria bearing an empty vector (-) or expressing wild-type or mutated AIDA-I proteins were tested for several phenotypes. *A*. Auto-aggregation assay. Cultures were left standing at 4°C and the OD<sub>600nm</sub> at the top of the culture was measured at the beginning of the assay (light gray) and after 180 minutes (dark gray). *B*. Biofilm formation assay. Biofilms formed in microtiter plates were stained with crystal violet and the amount of fixed dye measured by its absorption at 595 nm. *C*. Adhesion and invasion assays. Bacteria that adhered to Hep-2 cells were directly plated and counted. The number of intracellular bacteria was estimated by assessing the resistance to external antibiotics. The results represent the percentage of adhesion (light gray) or invasion (dark gray), compared to bacteria expressing wild-type AIDA-I. Significant differences ( $p < 0.05$ ) are indicated by stars for aggregation and biofilm formation (*A*, *B*), open circles for adhesion and dark circles for invasion (*C*). The difference in biofilm formation mediated by I4 was not statistically significant (NS).

### Fig. 3: Schematic representation of the insertion mutants in AIDA-I and their related phenotypes

Schematic representation of the location and effects of the mutations obtained from the linker scanning mutagenesis. White arrows correspond to the insertion mutants without

functional defect, whereas black arrows indicate mutants with a defect for one or more functions. The black box represents the HisG tag at the N terminus of mature AIDA-I.

**Fig. 4: Deletion of putative cell-binding domain in mature AIDA-I**

*A.* Schematic illustration of the deletion construct, representing plasmids pAngH (WT) and pAngH $\Delta$ N ( $\Delta$ N). The black box represents the HisG tag at the N terminus of mature AIDA-I. *B.* Whole cell lysates of bacteria harboring an empty vector (-), expressing wild-type AIDA-I (WT), or the  $\Delta$ N mutant were obtained and probed with anti-HisG antibodies, allowing detection of the proprotein (circle) and mature AIDA-I (arrowhead). All plasmids were transformed in bacteria expressing the Aah glycosyltransferase. *C.* Heat extracts of the same cultures were obtained, resolved by SDS-PAGE and stained with Coomassie blue. *D.* Detection of glycosylation, using the DIG glycan detection kit, performed on the whole cell lysates.

**Fig. 5: Function of AIDA-I bearing the  $\Delta$ N deletion**

Auto-aggregation assay (*A*), Biofilm formation (*B*), adhesion assay (*C*) and invasion assay (*D*) performed with cultures of C600 bearing an empty vector (-), plasmids pAngH and pAah (WT), or pAngH $\Delta$ N and pAah ( $\Delta$ N). The assays were performed as described in Fig. 2.

**Fig. 6: Cell-binding properties of glycosylated and unglycosylated GST-AIDAN**

*A.* Coomassie-blue-stained SDS-PAGE gel of the purification product from the soluble fraction of BL21 strain bearing plasmid pGex-4T-1, pGex-AIDAN or pGex-AIDAN and pAah using affinity chromatography with a glutathione sepharose column. The black arrows represent the position of the purified protein. *B.* Detection of glycosylation, using the DIG glycan detection kit, of the purified proteins. *C.* Cell-ELISA. Hep-2 cells were grown in 96-wells plate and after fixation and blocking, incubated with various concentrations of the whole AIDA-I (squares), GST (crosses), unglycosylated GST-AIDAN (filled circles) or glycosylated GST-AIDAN (open circles). Bound proteins were detected with an antibody against GST or AIDA-I coupled to HRP. The activity of HRP was measured using a colorimetric substrate by following absorption at 450 nm. Background was subtracted and the absorption was normalized by dividing the absorption

of each well by the maximal absorption measured on the plate in order to obtain percentages of maximal binding.

**Fig. 7: Immunofluorescent detection of purified unglycosylated and glycosylated GST-AIDAN on the surface of epithelial cells**

Purified GST, unglycosylated GST-AIDAN or glycosylated GST-AIDAN fusion proteins were added to confluent Hep-2 cells at a concentration of 1  $\mu$ M and incubated at 37°C for one hour. After binding of the proteins, the cells were washed, fixed and blocked with BSA before detection of bound proteins with antibody raised against GST. Immune complexes were revealed by incubation with a donkey anti-goat IgG rhodamine red conjugate.

**Fig. 8: Location of the 5 amino acids insertions in AIDA-I**

*A.* Sequences of the N-terminal part of the passenger domain of AIDA-I showing the imperfectly repeated 19-amino acid motif repeated 35 times. The positions of the insertion mutants are indicated. The circles highlight the positions of the identified glycosylation sites (M.-E. Charbonneau *et al.*, submitted manuscript). *B.* Weblogo (<http://weblogo.berkeley.edu/>) representation of the repeated motif showing the relative conservation of each amino acids of the consensus. Each repeat of this motif is thought to correspond to a coil of a  $\beta$  helix and is predicted to consist in 3  $\beta$  strands, as previously modelled (13). The positions of the strands, with the original nomenclature of the model, are indicated. *C.* Location of the 5-amino acid insertions on the coil model. A black circle represents an isoleucine, valine or leucine residue and “X” denotes any residue. White squares represent insertions that did not affect the functions of AIDA-I. Gray squares highlight insertions that disrupted one or more of these functions.

**Table I: Primers used in this study**

Primer	Sequence
$\Delta N$	5'-GGCCATCATCATCATCATCATGGTGCTAATGCCACTGAGACCATTGTCAATTC C -3'
$\Delta C$	5'- GGTAACCAGTATATTTCTGCTGGCGCAACCGTTACCGGAACTAACC -3'
$\Delta C2$	5'- GGTAACCAGTATATTTCTGCTGGCGCTGTTACATCCGGCGGCACAGCAACC- 3'
$\Delta NC$	5'- GGCCATCATCATCATCATCATGGTGCTGTTACATCCGGCGGCACAGCAACC- 3'
GST AidAN-F	5'- TATATAGGATCCGTAATATTTTCAGGCACAGTATCTTCAGG -3'
GST AidAN-R	5'- TATATACTCGAGTTAGCCAGCAGAAATATACTGGTTACC -3'

Figure 1

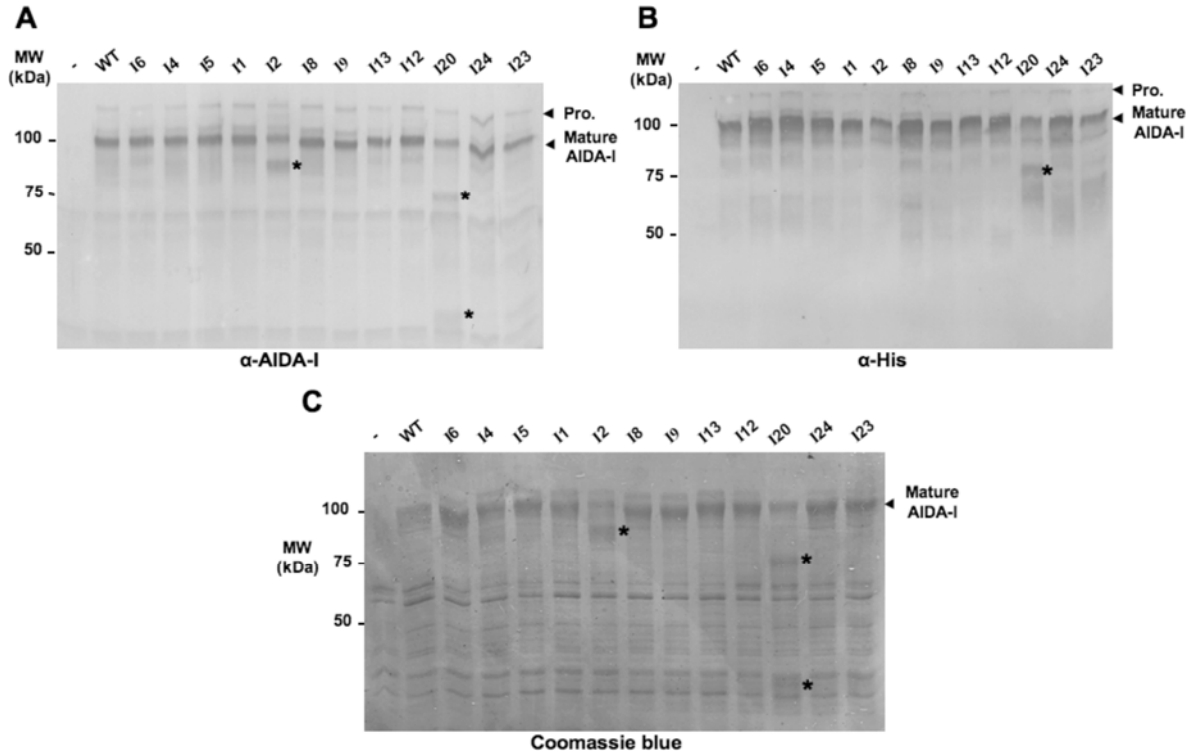


Figure 2

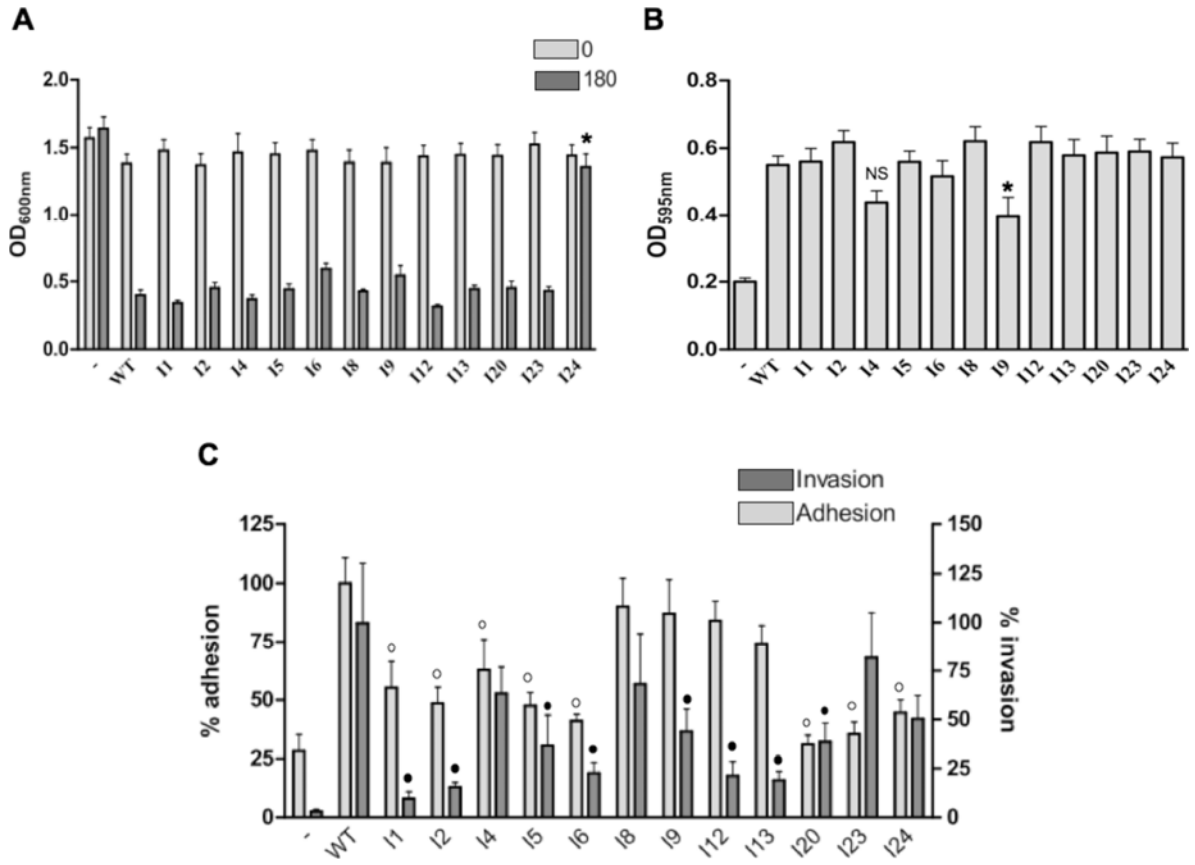




Figure 4

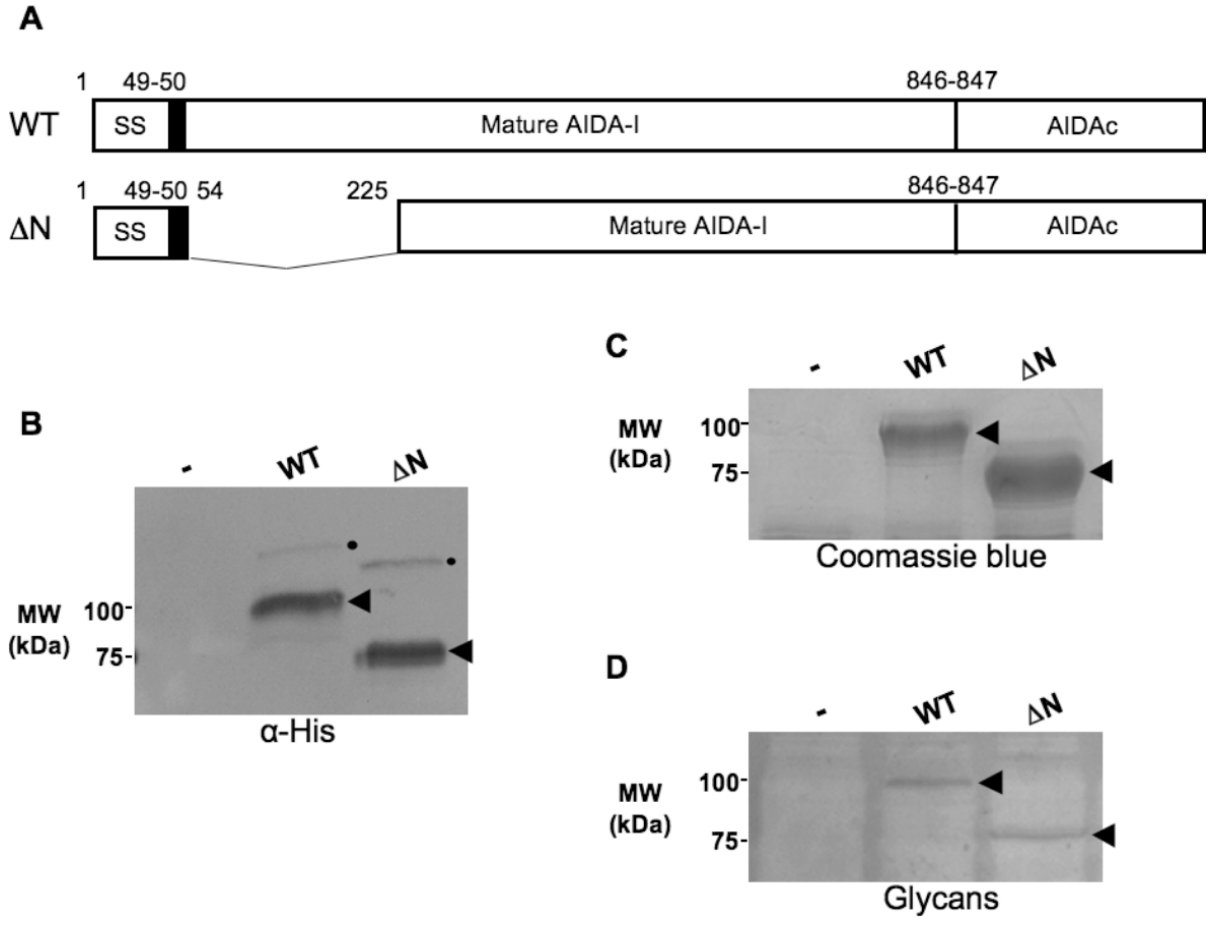




Figure 5

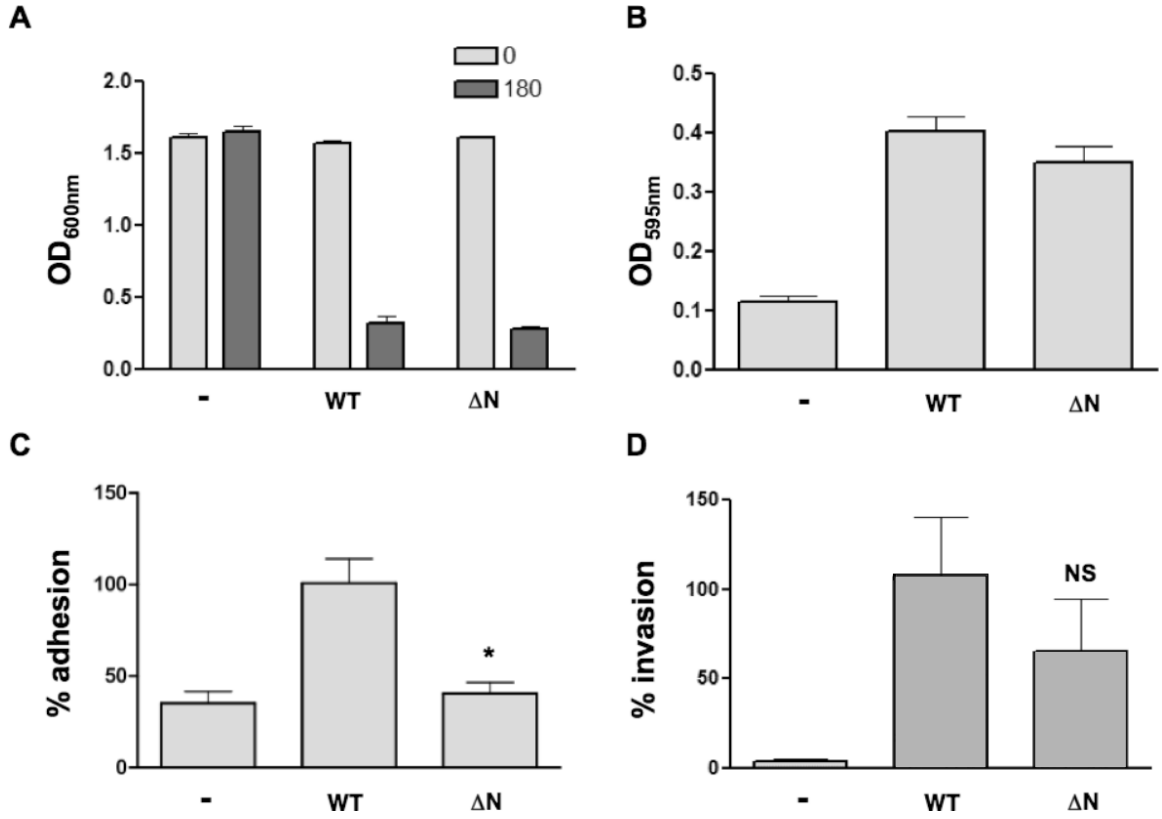


Figure 6

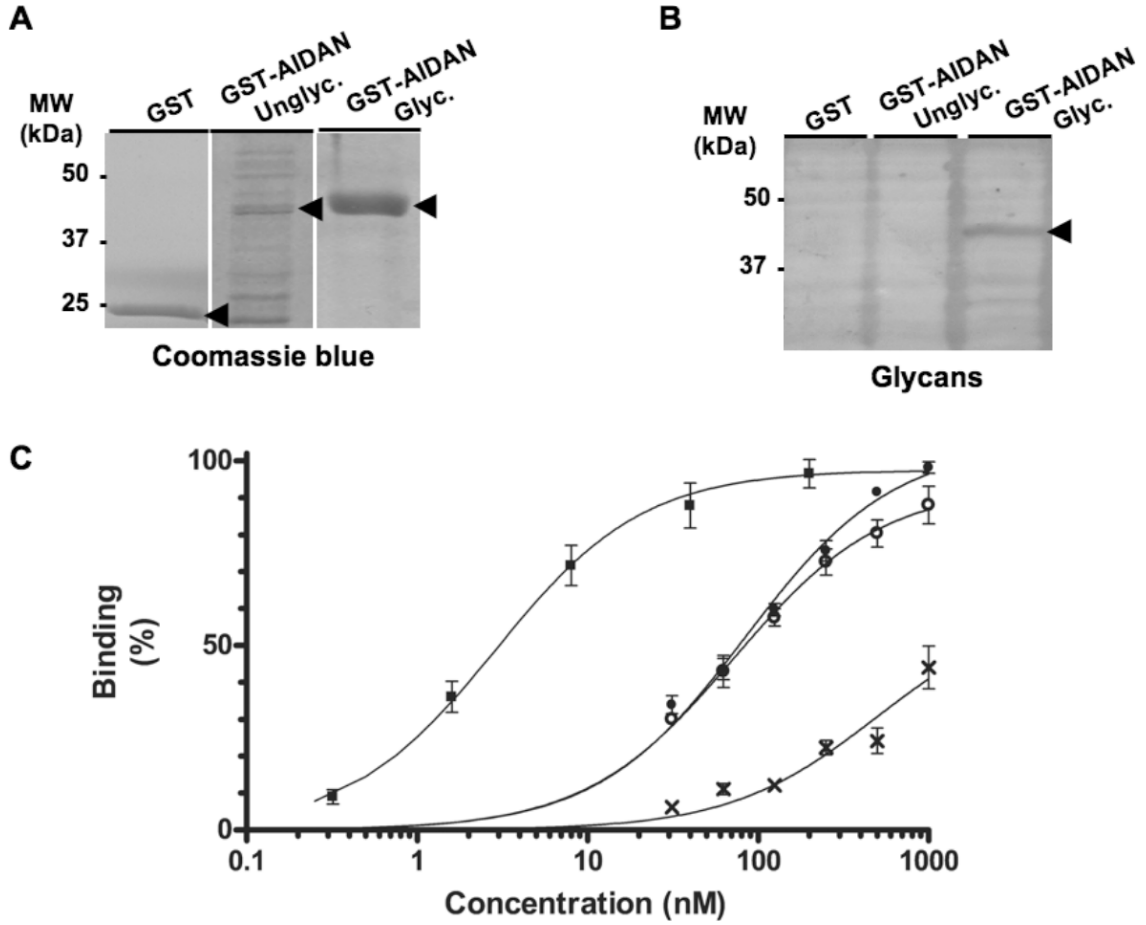


Figure 7

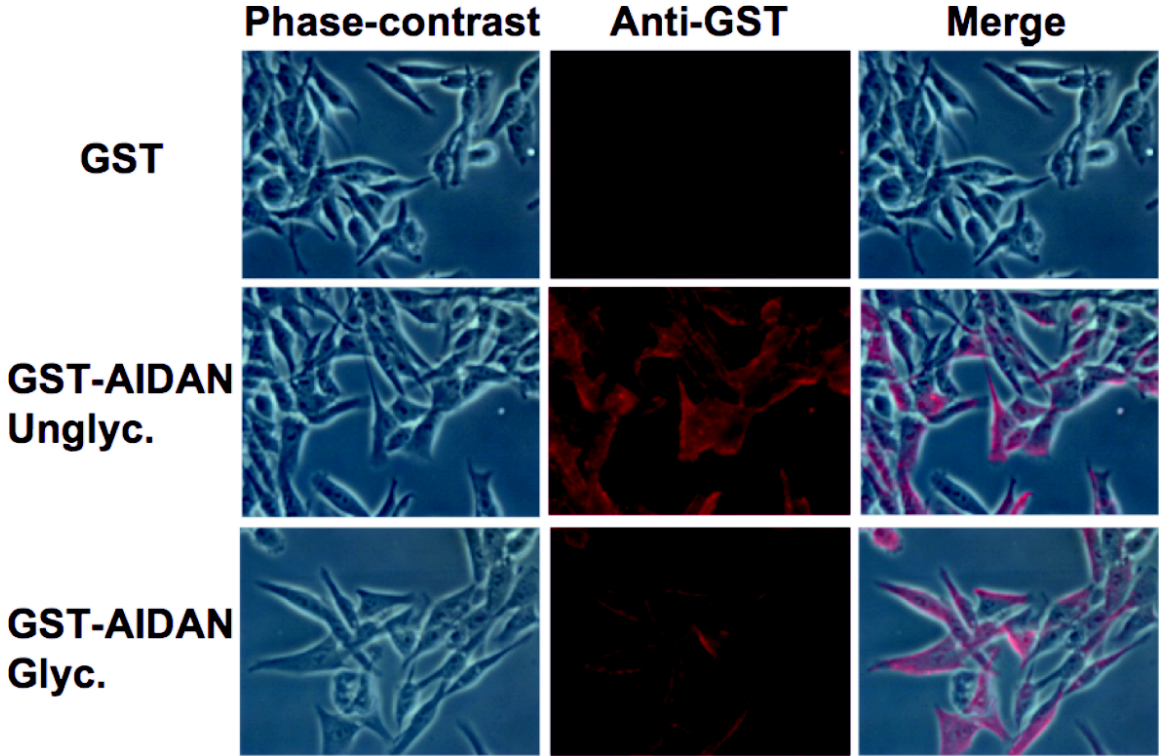
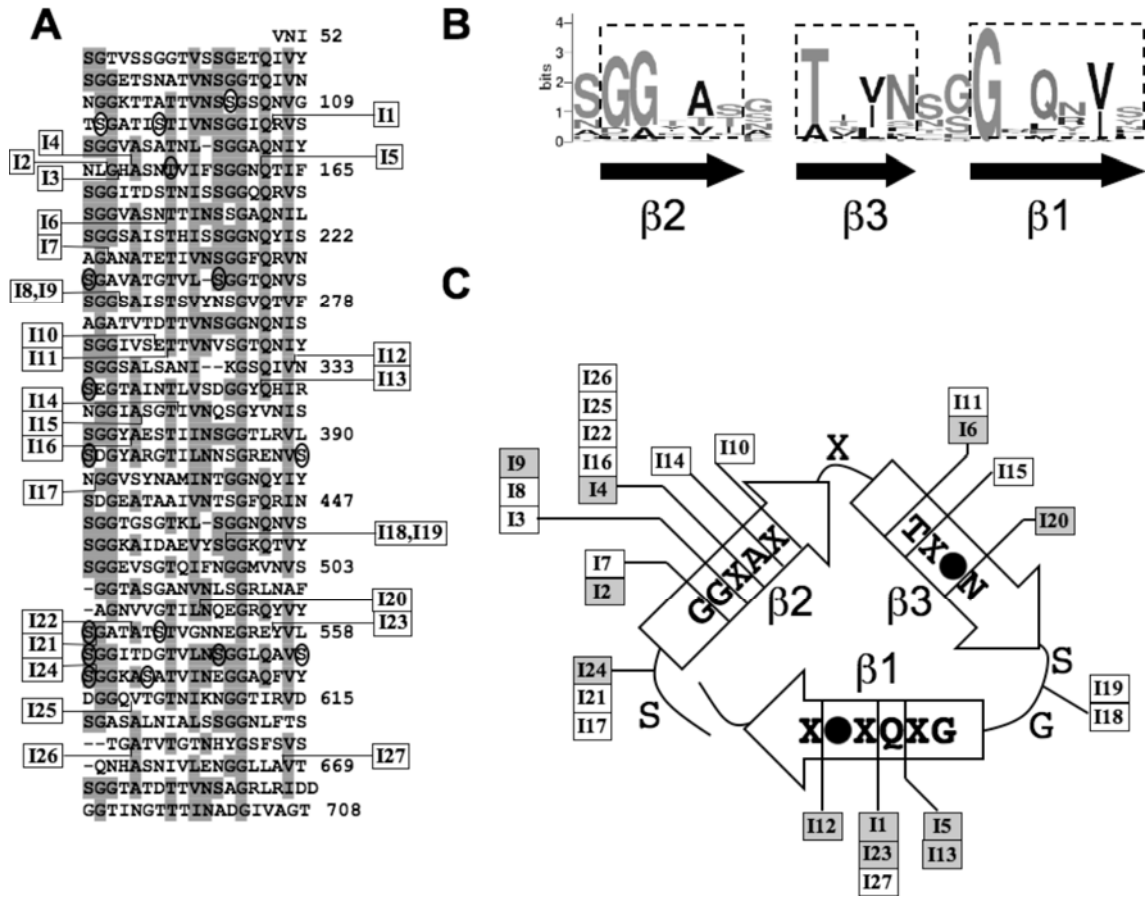


Figure 8



## ARTICLE 3

**O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence**

Journal of Bacteriology, 2007, Dec., Volume 189 (24), pages 8880-9

## Contribution de l'auteur pour cet article

Je suis co-premier auteur sur cet article. J'ai effectué les expérimentations concernant la fonctionnalité associée à l'expression de AIDA-I ainsi que celles qui démontrent que AIDA-I non-glycosylé est moins stable que AIDA-I glycosylé et que son expression induit un stress extracytoplasmique. De plus, à titre de co-premier auteur, j'ai participé activement à la rédaction de l'article.

## Contribution scientifique de l'article

Cet article montre que Aah est une O-glycosyltransférase qui modifie de façon hétérogène AIDA-I au niveau de résidus sérine ou thréonine. De plus, nous avons démontré que le rôle principal associé à la glycosylation est de maintenir AIDA-I dans une conformation normale, plutôt que de permettre une interaction directe avec les cellules hôte.

**O-linked glycosylation ensures the normal conformation of the autotransporter  
Adhesin Involved in Diffuse Adherence (AIDA-I)**

Marie-Ève Charbonneau<sup>1,†</sup>, Victoria Girard<sup>1,†</sup>, Anastasia Nikolakakis<sup>2</sup>, Manuel Campos<sup>1</sup>,  
Frédéric Berthiaume<sup>1</sup>, France Dumas<sup>3</sup>, François Lépine<sup>2</sup> and Michael Mourez<sup>1,\*</sup>

Running title: Glycosylation of the AIDA-I autotransporter

<sup>1</sup> Canada Research Chair on Bacterial Animal Diseases, Université de Montréal, Faculté de Médecine Vétérinaire, St-Hyacinthe, 3200 Sicotte, St-Hyacinthe, Québec, J2S 7C6, Canada. <sup>2</sup> INRS-Institut Armand-Frappier, Université du Québec, 531 Boul. des Prairies, Laval, Québec, H7V 1B7, Canada. <sup>3</sup> National Research Council, Biotechnology Research Institute, Publication #NRC47555, 6100 Royalmount, Montréal, Québec, H4P 2R2, Canada.

<sup>†</sup> Both authors contributed equally to this work

Keywords: Autotransporter, Bacterial Protein Glycosylation, Bacterial adhesin, AIDA-I, *Escherichia coli*

**Abstract**

The *Escherichia coli* Adhesin Involved in Diffuse Adherence (AIDA-I) is one of the few glycosylated proteins found in *Escherichia coli*. Glycosylation is mediated by a specific heptosyltransferase encoded by the *aah* gene but little is known about the role of this modification and the mechanism involved. In this study, we identified several peptides of AIDA-I modified by the addition of heptoses using mass spectrometry and N-terminal sequencing of proteolytic fragments of AIDA-I. One threonine and 15 serine residues were identified as bearing heptoses, thus demonstrating for the first time that AIDA-I is O-glycosylated. We observed that unglycosylated AIDA-I is expressed in smaller amounts than its glycosylated counterpart and shows extensive signs of degradation upon heat-extraction. We also observed that the unglycosylated AIDA-I is more sensitive to proteases and induces an important extracytoplasmic stress. Lastly, as was previously shown, we noted that glycosylation is required for AIDA-I to mediate adhesion to cultured epithelial cells, but purified mature AIDA-I fused to GST was found to bind *in vitro* to cells whether or not it is glycosylated. Taken together, our results suggest that glycosylation is required to ensure a normal conformation of AIDA-I and may be only indirectly necessary for its cell-binding function.

## INTRODUCTION

Protein glycosylation is widespread in prokaryotes, with more than 70 bacterial glycoproteins reported so far (38). Most of these are surface or secreted proteins that affect how bacteria interact with their environment, for instance by influencing cell-cell interactions, surface adhesion or evasion of immune response (34, 38).

In several bacterial species, complex O- and N-glycosylation pathways are encoded by multiples genes clustered in “glycosylation islands” (38, 39). In other cases, highly specific single glycosyltransferases are responsible for the modification of target proteins (2, 20, 35). However, in most cases the exact mechanism of glycosylation remains to be elucidated. How glycosylation exerts its role is also unclear. The addition of sugar moieties can define (9, 18) or mask (21, 31) interaction sites. Carbohydrates have also been shown to influence the stability (7) and protease sensitivity (14) of individual proteins or macromolecular assembly of several polypeptides (21, 33).

Only a few glycoproteins have been identified in *Escherichia coli*. Amongst them are the Adhesin Involved in Diffuse Adherence (AIDA-I) (2), the TibA adhesion/invasion protein of enterotoxigenic *E. coli* (20) and the auto-aggregation factor antigen 43 (Ag43) (35). Specific glycosyltransferases of AIDA-I and TibA have been identified but, by contrast, no Ag43-specific glycosyltransferases are known and Ag43 glycosylation was performed heterologously with the AIDA-I or TibA-specific enzymes. These glycoproteins present several similarities: all three (i) are secreted as autotransporters, a branch of the type V secretion pathway; (ii) have nearly identical N-terminal 19 amino acids repeats; (iii) are glycosylated by the addition of heptoses mediated by single glycosyltransferases that are functionally interchangeable; and (iv) are versatile virulence factors mediating bacterial auto-aggregation, biofilm formation, as well as adhesion to and invasion of mammalian cells. Because of these similarities, AIDA-I, TibA and Ag43 have been named Self-Associating Autotransporters (SAAT) (17).

AIDA-I was originally identified as a plasmid-encoded protein able to confer a pattern of diffuse adherence on the surface of cultured epithelial cells (1). It is associated with a high percentage of the pathogenic *E. coli* strains involved in neonatal and post-weaning diarrhea in piglets, which cause major economic losses in farms worldwide (10, 11, 22, 26, 27). AIDA-I is synthesized as a 132 kDa pre-proprotein (37). A cleavable N-terminal signal sequence of 49 amino acids allows secretion across the inner membrane via the general *sec* secretion machinery. After crossing the periplasm, the proprotein is inserted



in the outer membrane and is cleaved, probably by an autocatalytic mechanism (3, 37). The cleavage separates an N-terminal extracellular fragment, the mature AIDA-I adhesin, from a C-terminal membrane-embedded fragment, AIDAc. Despite the cleavage, the mature adhesin and AIDAc remain strongly associated at the bacterial surface (3, 37).

The autotransporter adhesin heptosyltransferase (Aah) is responsible for the glycosylation of AIDA-I with heptoses. Aah uses precursors recruited from the lipopolysaccharide biosynthetic pathway, but the modified residues were not identified (2). The glycosylation of AIDA-I was shown to be essential for adhesion, since deletion of the *aah* gene abolishes adherence to cultured epithelial cells, but dispensable for auto-aggregation and biofilm formation (2, 36). These observations suggested that heptose residues are involved in the interaction between AIDA-I and a receptor on the surface of epithelial cells. There is however no formal proof for this hypothesis and glycosylation could also affect the conformation or the stability of the protein. It was for instance shown that glycosylation was required for the stability of the HMW1 adhesin of *Haemophilus influenzae* and its tethering to the bacterial surface (7).

In the present study, we characterized the glycosylation of AIDA-I. We used mass spectrometry (MS) and N-terminal sequencing of AIDA-I peptides to show that the heptoses are O-linked to the 15 serine and one threonine residues. We also show that glycosylation provides increased resistance to degradation and does not alter the binding efficiency of purified mature AIDA-I. Our results therefore strongly suggest that glycosylation is mainly necessary for the stability of the protein at the cell surface rather than being required to engage a cellular receptor.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** The *Escherichia coli* K12 strain C600 (*thr-1*, *leuB6*, *thi-1*, *lacY1*, *supE44*, *rfbD1*, *fhuA21*, obtained from New England Biolabs) and the pathogenic strain 2787 (1) were used in this study. Plasmid pAg containing the whole *aidA* operon (*aah* and *aidA*) under the control of *ptrc*, a promoter inducible with isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG), has been described before (3). pAg allows the expression of glycosylated AIDA-I. Plasmid pAgH, also previously described (3), is derived from pAg and allows expression of glycosylated AIDA-I tagged at the N terminus of the proprotein with 6 histidine amino acids and a glycine (HisG). To construct plasmid pAng, which allows the expression of non-glycosylated AIDA-I, we amplified by PCR the *aidA* gene

from the *E. coli* strain 2787 using primers introducing *NcoI* and *XbaI* restriction sites. The resulting fragment was cloned in the vector pTRC99a (Pharmacia Biotech), resulting in plasmid pAng that bears the *aida* gene alone under the control of the *ptrc* promoter. Plasmid pAngH, allowing the expression of the HisG-tagged AIDA-I, was obtained by introducing an oligonucleotide coding for the HisG tag in the pAng plasmid by site directed mutagenesis using mutagenic primers, as described previously with pAg (3). To construct the plasmid pAah, we subcloned the *aah* gene with its *ptrc* promoter from the pTRC-Aah plasmid described before into the plasmid pACYC184 (Pharmacia Biotech) using *SphI* and *EcoRI* restriction sites. The pAah plasmid has an origin of replication compatible with those of pAng and pAngH. A plasmid allowing the expression of a fusion between Glutathione-S-transferase (GST) and mature AIDA (encompassing residues 50 to 847) was constructed by PCR amplification and cloning into the pGex-4T-1 vector (Amersham biosciences), as described elsewhere (M.-E. Charbonneau and M. Mourez, submitted for publication). All constructions were verified by restriction mapping and sequencing.

**Bacterial and cell culture growth conditions.** Bacteria containing the different plasmids were grown on Luria-Bertani (LB) agar plate or in liquid LB (rich) or M9 (minimal) media containing 100  $\mu\text{g}\cdot\text{ml}^{-1}$  of ampicillin, and, in addition, 50  $\mu\text{g}\cdot\text{ml}^{-1}$  chloramphenicol when the plasmid pAah was used. Bacterial cultures were grown at 30°C, and induced overnight when they reached an  $\text{OD}_{600\text{nm}}$  of 0.8 with 10  $\mu\text{M}$  of IPTG, unless indicated otherwise. Strain 2787 was grown on I-medium (1) at 37°C. Hep-2 cells (ATCC CCL-23) were grown at 37°C with 5%  $\text{CO}_2$  in Dulbecco's modified Eagle medium (Gibco) containing 10 mM sodium pyruvate (Sigma), bovine growth serum (Hyclone), 2.5  $\mu\text{g}\cdot\text{ml}^{-1}$  of fungizone and 100  $\mu\text{g}\cdot\text{ml}^{-1}$  of penicillin/streptomycin (Gibco).

**SDS-PAGE and immunoblotting.** Protein-containing samples were diluted in twice concentrated SDS-PAGE loading buffer containing  $\beta$ -mercaptoethanol and denatured by heating at 100°C for 10 minutes. The samples were separated by SDS-PAGE on 10% acrylamide gels. The gels were either stained with Coomassie blue or transferred to polyvinylidene fluoride membranes (Millipore). Immunodetection was performed with an anti-HisG horseradish peroxidase (HRP)-coupled antibody (Invitrogen) diluted 1:5,000 in blocking buffer (5% skim milk, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Triton X100). Immune complexes were revealed using a 3,3',5,5'-tetramethylbenzidine solution

for membranes (Sigma). For quantitative comparison the membranes were digitally scanned using the ImageJ software (NIH, USA).

**Whole cell and membrane extract preparations.** Overnight cultures from *E. coli* C600 harboring pTRC99A, pAngH (non-glycosylated AIDA-I) or pAngH and pAah (glycosylated AIDA-I) normalized to the same OD<sub>600nm</sub>. To obtain whole cell extracts, the cultures were normalized, pelleted and bacteria resuspended in Phosphate buffered saline (PBS). To obtain membrane fractions, the cultures were lysed and fractionated as described previously (3).

**Heat extraction and purification of mature or whole AIDA-I.** Heat extraction was performed as described before (3). Briefly, overnight cultures of C600 harboring empty vector, pAngH or pAngH and pAah were normalized and bacteria were harvested, resuspended in 10 mM sodium phosphate buffer, pH 7 and heated at 60°C for 20 minutes. The treated samples were centrifuged for 5 minutes at 12,000 x g to recover the heat-extracts. The HisG-tagged mature AIDA-I was purified from heat extracts and the HisG-tagged whole AIDA-I was purified from solubilized membrane extracts, as described before (3).

**Protease accessibility assay.** Normalized cultures of C600 bearing an empty vector, pAngH or pAngH and pAah were pelleted and resuspended in PBS in the presence or absence of trypsin (Sigma) or proteinase K (Invitrogen), at 0.3 µg.ml<sup>-1</sup> or 3 mg.ml<sup>-1</sup>, final concentrations. After 30 minutes of incubation on ice, the proteases were neutralized by addition of proteases inhibitor cocktail (Complete mini; Roche) for 5 minutes. The samples were pelleted, and whole cell extracts obtained as described above.

**Purification of the GST fusion proteins.** One liter of *E. coli* BL21 harboring plasmids pGex-4T-1 or pGex-AIDA with or without pAah were grown until an OD of 0.4 and induced with 10 µM of IPTG for three hours. Bacteria were harvested and resuspended in 40 ml of Tris-buffered saline (50 mM Tris-HCl pH 8, 150 mM NaCl) containing lysozyme (0.4 mg.ml<sup>-1</sup> final) and EDTA pH 8 (10 mM final) and lysed with a French press and an ultrasonic processor. The soluble fraction was isolated by a 30-minute centrifugation at 16,000 x g. The GST and the GST-AIDA fusion proteins were purified using an ÄKTA

purifier system with a 5-ml glutathione sepharose column (Amersham biosciences), according to instructions of the manufacturer. The purity of the purified proteins was confirmed by SDS-PAGE and staining with Coomassie blue. Glycan detection was performed using the DIG glycan detection kit (Roche), according to the instructions of the manufacturer.

**Cell-ELISA.** Hep-2 cells were grown in 96-wells plate, and fixed 15 minutes with PBS containing 2.5% paraformaldehyde and 0.2% glutaraldehyde as final concentrations. After one wash with PBS, the plate was blocked with PBS-BSA 3% for 2h at 37°C. Purified proteins (GST, a fusion of GST to mature AIDA-I, or whole AIDA-I) were added to the cells at concentrations between 30 nM and 1 $\mu$ M and incubated overnight at 4°C. After extensive washes with PBS, bound proteins were detected with an anti-GST antibody coupled to HRP, diluted 1:10,000 in PBS (Amersham biosciences), or with a custom polyclonal rabbit anti-AIDA serum (QCB, Hopkinton, MA) and a secondary goat anti-rabbit antibody coupled to HRP. Immune complexes were revealed using a 3,3',5,5'-tetramethylbenzidine solution for ELISA (Sigma). Background was subtracted and the absorption intensities were normalized by dividing the intensity in each well by the maximum intensity measured on the plate. Experiments were conducted in duplicate at least twice. Binding curves and dissociation constants were obtained by nonlinear regression fitting to a one binding site hyperbola, using Prism 4.0 (Graphpad software).

**Functional assays.** All assays were performed as described before (3). Briefly, the auto-aggregation assay was performed using overnight cultures of C600 harboring empty vector, pAng or pAg, or cultures of strain 2787. All cultures were normalized, vortexed for 10 seconds and left at 4°C. The OD<sub>600nm</sub> was measured at the top of the culture at the beginning of the assay and after 3h, and pictures were taken after overnight incubation. For biofilm formation, normalized cultures of C600 harboring empty vector, pAng or pAg were grown without agitation for 24 h at 30°C in minimal medium in 96-well PVC plates (Falcon). Biofilms were stained for 15 minutes with crystal violet 1% and the fixed dye was solubilized by addition of ethanol-acetone (80:20). The absorbance of the dye solution was measured at 595nm. For unknown reasons strain 2787 grows extremely slowly in minimal medium and AIDA-I expression is minimal in those conditions, the biofilm assay was therefore not performed with this strain. The adhesion assay was performed with Hep-2

cells, grown in 24-well plates, inoculated with  $10^6$  cfu per well of C600 harboring empty vector, pAng or pAg for 3 h. Cells were washed with PBS. Bacteria adhering to cells were recovered with 100 ml of Triton X100 1% and plated for numbering. Adherence was calculated by dividing the number of adherent bacteria with the number of bacteria found in the inoculum after three hours of incubation.

**$\beta$ -galactosidase reporter assay.** The  $\beta$ -galactosidase activity of strains SR1458 (30) and SR1364 (25) transformed with an empty vector, plasmid pAng or plasmid pAg was assessed as described previously (25) and expressed in Miller units. Statistical comparisons were performed by variance analysis (ANOVA) using Prism 4.0 (Graphpad software).

**Mass spectrometry.** Glycosylated AIDA-I was purified from heat-extracts, as described above, and run on a SDS-PAGE 10% acrylamide gel. The protein band corresponding to AIDA-I was cut from the gel, destained with water/sodium bicarbonate buffer and acetonitrile. The protein was reduced with DTT and alkylated with iodoacetamide prior to in-gel digestion with trypsin or chymotrypsin. The tryptic peptides were eluted from the gel with acetonitrile containing 0.1% of trifluoroacetic acid. The tryptic peptides were then separated on an Agilent Nanopump using a C18 ZORBAX trap and a SB-C18 ZORBAX 300 reversed phase column (150 mm x 75  $\mu$ m, 3.5 $\mu$ m particle size) (Agilent Technologies, Inc.). All mass spectra were recorded on a hybrid linear ion trap-triple quadrupole mass spectrometer (Q-Trap, AB Applied Biosystems, MDS SCIEX Instruments, California, USA) equipped with a nano-electrospray ionization source. The accumulation of MS-MS data was performed with the Analyst Software, version 1.4 (AB Applied Biosystems / MDS SCIEX Instruments, California, USA). MASCOT (Matrix Science, London, UK) was used to create peak lists from MS and MS/MS raw data.

**N-terminal sequencing.** The mature AIDA-I protein was isolated from a SDS-gel stained with Coomassie Blue as described above. The band corresponding to mature AIDA-I was cut and transferred in a 1.5 ml screw cap microfuge tube, reduced, alkylated and digested with trypsin (Promega sequencing grade) as described previously (12). The peptides were extracted with 60% acetonitrile, 1% trifluoroacetic acid at 60°C and separated by reverse-phase HPLC on a Vydac microbore C18 column (1mmIDx50mm) using an Applied Biosystems 130A Separation System. The peptides were detected by absorbance at 220 nm

and fractions peaks were collected manually. Each fraction was applied to a precycled glass fiber filter treated with trifluoroacetic acid and coated with Biobrene Plus (0.5 mg of polybrene and 0.03 mg of NaCl). The fraction were subjected to automatic Edman degradation on a model 494 CLC Procise sequencer using a general protocol (15). The phenylthiohydantoin amino acid derivatives were analyzed on-line using a capillary separation system (Applied Biosystems model 140 D) and an ultraviolet detector (Applied Biosystems model 785A) set at 269 nm.

## RESULTS

**AIDA-I is O-glycosylated.** We purified glycosylated mature AIDA-I (corresponding to amino acids 50 to 847 in the pre-proprotein) by affinity chromatography from heat-extracts of the *E. coli* strain C600 expressing a histidine-tagged protein, as described before (3). After SDS-PAGE, the band corresponding to purified mature glycosylated AIDA-I was in-gel digested by trypsin or chymotrypsin and the resulting peptides were submitted to tandem MS/MS with collision induced dissociation (CID). Some of the parent ions were identified as AIDA-I peptides presenting between 1 and 4 incremental additions of 192 Da compared with the theoretical mass of the unsubstituted peptides. This 192 Da addition is consistent with the grafting of a heptose molecule and thus we concluded that these peptides were glycosylated. In most cases we could compare the MS spectra of modified and unmodified peptides. With this MS approach, nine peptides were found to be glycosylated with 1 to 4 heptose residues whereas five peptides were found not to be glycosylated (Fig. 1 and Table 1). In two instances, there was extensive overlap between two glycosylated peptides, therefore we in fact identified seven different glycosylated peptides using MS. We noted that the glycosylation is heterogeneous, as five out of the nine peptides were found without carbohydrate or with various amounts of heptose residues. When multiple glycosylations was observed, it was not possible to determine if it was the result of multiple sites being glycosylated or a single site being glycosylated by multiple heptose residues. Similarly, we could not ascertain the conformation of the grafted heptose. The glycosylated peptides are found at several different positions in mature AIDA-I, but exclusively in a region containing 35 imperfect repeats of a 19 amino acids sequence (Fig. 1). In most cases, CID of peptides does not allow the identification of the amino acid residue bearing the carbohydrate since the carbohydrate-polypeptide bond is usually preferentially cleaved compared to the peptide bond. However, in one instance, the CID

spectrum of a chymotryptic peptide showed peptide fragment ions still bearing the carbohydrate moiety (Fig. S1). Analysis of these fragment ions revealed that the threonine at position 154 was modified with one heptose.

In a parallel approach, purified tryptic peptides of mature AIDA-I were sequenced by N-terminal Edman degradation (ED). Six peptides had the expected sequences of AIDA-I peptides, with perfect signal for each residue of the peptides. However, for 8 other peptides, no signal was observed at positions corresponding to some of the serine residues they comprised (Table 1). This suggested that these residues had been modified and thus could not be identified during the degradation cycle. Five of the eight peptides identified as potentially modified corresponded to peptides identified by MS as glycosylated, suggesting that all the peptides identified in this manner by ED were indeed glycosylated. The signal corresponding to the serine residues at positions 577 and 578 was sharply reduced but not absent, suggesting that these residues are heterogeneously modified. This result is in agreement with the heterogeneity observed by MS. Also in agreement with the previous results, the modified peptides identified by ED are located in the 19-amino acids repeats.

Together, our approaches revealed ten different glycosylated peptides and 15 serine and one threonine residues were found to be modified. Adding the maximum number of heptoses bound to each of the peptide identified by MS and the number of modified serine residues in the peptides identified solely by ED, we estimate that up to a total of 19 heptoses could be present on AIDA-I (Table 1). All attempts to obtain the mass of the whole mature AIDA-I by MS were unsuccessful. This is most likely due to the high molecular weight of the protein and the heterogeneity conferred by the glycosylation.

**Glycosylation influences the abundance of AIDA-I.** We compared the ability of bacteria expressing the glycosylated or the non-glycosylated forms of AIDA-I to auto-aggregate, form a biofilm or adhere onto cultured epithelial cells. Glycosylation seems dispensable for auto-aggregation and biofilm formation, but the non-glycosylated form of AIDA-I cannot mediate adhesion to Hep-2 cells (Fig. 2), as previously observed (2, 36). This observation could be due to the fact that unglycosylated AIDA-I is unable to mediate adhesion, that there is not enough unglycosylated AIDA-I present at the bacterial surface to mediate adhesion, or that the conformation of unglycosylated AIDA-I is abnormal and cannot mediate adhesion.

To distinguish between these possibilities, we compared the amounts of the glycosylated and non-glycosylated forms of AIDA-I present at the cell surface. As previously suggested, we observed that antibodies directed against the glycosylated mature adhesin do not efficiently recognize non-glycosylated AIDA-I (2), hindering efforts to compare expression levels. To remedy this situation, we used a plasmid allowing the expression of non-glycosylated AIDA-I with a HisG-tag localized at the N terminus of the mature adhesin. In order to express the glycosylated protein, we co-transformed bacteria with this plasmid and a compatible plasmid containing the *aah* gene. With those constructs, we could prepare whole cell extracts of normalized overnight cultures expressing glycosylated or unglycosylated forms of the histidine-tagged protein. Proteins of 100 and 132 kDa for glycosylated AIDA-I and 80 and 120 kDa for non-glycosylated AIDA-I, were detected with an anti-HisG monoclonal antibody, consistent with the mature protein and the proprotein precursor forms of AIDA-I (Fig. 3A). We observed that there was dramatically less of the non-glycosylated form of AIDA-I compared to the glycosylated form. We observed the same decrease with untagged versions of AIDA-I by separating membrane fractions of bacteria expressing glycosylated or unglycosylated forms of AIDA-I and staining with Coomassie blue (data not shown).

Altered expression levels might be due to the fact that glycosylated AIDA-I is expressed in bacteria bearing two plasmids, whereas unglycosylated AIDA-I is expressed in bacteria bearing only one plasmid. To eliminate this possibility, we expressed unglycosylated AIDA-I in bacteria bearing one plasmid by engineering a point mutation in the *aah* gene, resulting in the expression of an inactive glycosyltransferase, as previously observed (24). As above, we observed dramatically less non-glycosylated AIDA-I compared to its glycosylated form in heat-extracts (data not shown).

**Glycosylation confers partial resistance to proteases.** We observed that when the mature unglycosylated and glycosylated adhesins were heat-extracted, the glycosylated polypeptide runs as a clear band of approximately 100 kDa, whereas the non-glycosylated protein appears as a less intense band of approximately 80 kDa, along with a degradation profile absent from extracts of bacteria expressing glycosylated AIDA-I (Fig. 3B). Additionally, the heat-extracted unglycosylated polypeptides were undetectable with antibodies directed against the HisG tag (data not shown). This result suggests that unglycosylated AIDA-I is more sensitive to proteolytic degradation. To determine if and



where unglycosylated AIDA-I is degraded *in vivo*, we compared the expression levels of glycosylated and unglycosylated AIDA-I in a *degP* background, but there was no visible effect of the presence of the periplasmic protease DegP on the relative amounts of unglycosylated or glycosylated AIDA-I (Fig. S2). We also tested if the outer membrane protease OmpT could have an influence on the levels of AIDA-I but it also seemed not to have an effect (Fig. S2). Since AIDA-I is auto-proteolytically maturing itself (3, 37), it could be that unglycosylated AIDA-I is responsible for its own degradation when it is in an abnormal conformation.

To confirm the hypothesis that unglycosylated AIDA-I is more sensitive to proteolytic degradation, we performed a limited digestion of bacterial surface proteins using trypsin, a protease cleaving after lysine and arginine residues. We observed that the unglycosylated AIDA-I is degraded at the lowest concentration of trypsin whereas the glycosylated protein is resistant at all concentrations tested (Fig. 4).

To ensure that this result is not specific for trypsin, which could simply indicate that most trypsin cleavage sites are protected by the heptoses, we performed the same assay with various concentrations of proteinase K, a protease cleaving after hydrophobic residues. Consistent with our previous results, we observed that the unglycosylated protein is completely degraded at the lowest concentration of proteinase K used (Fig. 4). To further exclude artifacts that might be due to the use of histidine-tagged proteins, we performed the same assays with untagged proteins, revealed by Coomassie blue staining. As above, in these experiments, the unglycosylated protein was completely degraded by  $0.3 \mu\text{g}\cdot\text{ml}^{-1}$  of proteinase K whereas the glycosylated protein was resistant (data not shown).

**The unglycosylated form of AIDA-I induces an extracytoplasmic stress.** The sensitivity of unglycosylated AIDA-I to proteolytic degradation suggests that it is in an abnormal conformation. Polypeptides with abnormal conformation are usually sensed by specialized stress sensing systems, which in turn induce the production of folding catalysts and proteolytic enzymes to alleviate the problem (23). We therefore tested if unglycosylated AIDA-I can induce a stress response. We used two reporter strains: strain SR1458, which bears the  $\beta$ -galactosidase *lacZ* gene under the control of the *degP* promoter (30) and strain SR1364, a strain bearing the *lacZ* gene under the control of the *rpoH* promoter (25). The former is induced by an extracytoplasmic stress and the latter by a cytoplasmic stress. The strains were transformed with the plasmids bearing the *aah-aidA* operon, the *aidA* gene

alone or a control empty plasmid (Fig. 5). We observed that glycosylated and unglycosylated AIDA-I did not induce any cytoplasmic stress (Fig. 5A). The  $\beta$ -galactosidase activity indicated however that unglycosylated AIDA-I caused a dramatic extracytoplasmic stress response (Fig. 5B). It is important to note that SR1458 is only a reporter strain and the induction of the *degP* promoter does not prove that unglycosylated AIDA-I is degraded by the periplasmic DegP protease. Indeed, as described above, we have observed that DegP does not seem to be involved in the degradation of unglycosylated AIDA-I. It is also important to note that these results were obtained in a context of overexpression and we do not know if a similar stress response would be induced in an *aah* mutant of the wild-type strain 2787. Nevertheless, the stress response induction upon overexpression are consistent with the enhanced sensitivity to proteolytic degradation and suggests that unglycosylated AIDA-I adopts an abnormal conformation.

**Glycosylation is not required for the binding of a domain of AIDA-I to cultured epithelial cells.** We decided to test directly the role of glycosylation on the binding of AIDA-I to epithelial cells by comparing the binding of glycosylated and unglycosylated AIDA-I. Whereas we can purify whole glycosylated AIDA-I by solubilizing outer membrane extracts (3), we are unable to purify unglycosylated AIDA-I because it is unstable and degraded. We therefore purified a fusion of mature AIDA-I (corresponding to amino acids 50 to 847) to GST, unglycosylated or glycosylated by co-expression in the presence of Aah. The presence or absence of glycosylation of the GST-AIDA was ascertained using a glycan detection kit (data not shown). Glycosylated GST-AIDA bound specifically to cultured epithelial cells in a saturable manner (Fig. 6). The equilibrium dissociation constant was calculated to be approximately 45 nM. By comparison, we observed that whole glycosylated AIDA-I could bind to cells with a dissociation constant of 4 nM, in close agreement with a previous estimate (19). In a structure-function study we have observed that the N terminus of AIDA-I harbors a cell-binding domain (M.-E. Charbonneau and M. Mourez, submitted for publication). In this context, when the GST is fused to the N terminus of mature AIDA-I it is likely that it will sterically hinder the interaction mediated by the N-terminal cell binding domain. This might explain the difference between the dissociation constant of our GST-fusion protein and that of wild-type AIDA-I. The fusion of mature AIDA to GST, purified in the absence of Aah (and therefore unglycosylated), was also able to bind to cultured epithelial cells in a manner

similar to the glycosylated fusion, with an equilibrium dissociation constant calculated at 39 nM. This result suggests that the carbohydrates moieties are not directly participating in the binding to a cellular receptor.

## DISCUSSION

In the present study, we identified ten peptides of AIDA-I bearing heptose residues and our results suggest that at least up to 19 molecules can be added onto AIDA-I, in perfect agreement with a previous evaluation (2). Recently, five different peptides were found to contain heptoses in Ag43 (35). Strikingly, in AIDA-I as well as in Ag43, glycosylation appears to be heterogeneous, i.e. most of the glycosylated peptides could be identified unglycosylated and/or with various numbers of heptoses. This fact was not initially appreciated when AIDA-I was shown to be glycosylated (2). The coexistence of modified and unmodified forms of bacterial glycoproteins has previously been reported (31, 41). In some instances it has been shown that this heterogeneity influences the immunogenicity of the protein (31), but the role, if any, of this heterogeneity in AIDA-I is unclear. The mechanism that results in such heterogeneity is also unknown.

The glycosylated peptides were identified in a region of the protein composed of imperfect 19-amino acids repeats. The same was true with Ag43 (35). It was expected that the specificity of the heptosyltransferases acting on these proteins is similar since the SAAT glycosyltransferases are functionally exchangeable (24, 35). The role of these repeats is unknown but it has been postulated that they are involved in the adhesion mediated by these proteins, since it is often observed that repeated motifs are involved in adhesion. Alternatively, the repeated motifs could provide the backbone for the expected  $\alpha$ -helical structure of the proteins, itself a repetitive structure.

In addition to what was reported with Ag43, we could unambiguously show that AIDA-I is modified on serine and threonine residues, proving that Aah mediates O-glycosylation. This was expected, based on sequence similarities of Aah with the known *E. coli* heptosyltransferases involved in LPS biosynthesis (2). Indeed, these enzymes mediate the transfer of heptose precursors onto the hydroxyl group of LPS biosynthetic intermediates (8). We did not identify any glycosylation consensus sequence but it should be noted that neither in eukaryotes, nor in prokaryotes, has a consensus sequence for O-glycosylation been established (29, 34, 38, 40). Consequently, the specificity of the Aah glycosyltransferases remains elusive: How does Aah recognize its SAAT substrates and

why does it transfer heptoses in the 19-amino acids imperfect repeats? It is possible that this region adopts a specific structure that is recognized by the enzyme but this would contradict the fact that glycosylation is thought to occur in the cytoplasm, prior to transport across the inner membrane via the *sec* machinery, which requires exported polypeptides to be unfolded. Alternatively, the enzyme might recognize a part of the repeated consensus sequence itself. Specificity of O-glycosylation towards amino acid sequences has been proposed before, as in the case of proline-rich domains of glycoproteins from *Mycobacterium tuberculosis* (4) and *Clostridium thermocellum* (6).

Important biological functions can be affected by glycosylation: maintenance of protein conformation, resistance against proteases, or modulation of intermolecular interactions (2, 38). Our results confirmed that glycosylation of AIDA-I is essential for adhesion (2), but not for auto-aggregation or biofilm formation (36). Based on these results, it was suggested that the glycans could be involved in receptor recognition. Carbohydrates from glycoproteins have indeed been shown in some cases to mediate the interaction with the host cell receptor, as for the major outer membrane protein of *Chlamydia trachomatis* (18). All our observations, however, suggest that glycosylation is required for AIDA-I to adopt a normal conformation, which in turn would be responsible for the lack of adhesion. Indeed, we observed that unglycosylated AIDA-I is less expressed and more sensitive to degradation but that purified mature AIDA-I could specifically bind to cultured epithelial cells even when unglycosylated. There are other examples of bacterial proteins being protected against degradation by glycosylation (14), and the conformation and protease sensitivities of polypeptides have often been shown to be altered by glycosylation (5, 16). The *H. influenzae* HMW1 adhesin, for instance, is stabilized by glycosylation (7). Unglycosylated HMW1 was prematurely degraded in the cytoplasm and periplasm and its tethering to the bacterial surface was compromised. Interestingly, HMW1 and SAAT are both secreted by the type V secretion pathway (13), raising the possibility that glycosylation might also exerts its role in the context of the secretion of some substrates of this pathway. Such a role could be to prevent premature periplasmic folding, since it was shown that it can be incompatible with secretion (32), or to promote extracellular folding, a process which has been proposed to drive secretion (28).

Two observations seem to be at odds with the notion that unglycosylated AIDA-I is in an abnormal conformation. First, Ag43 is usually not glycosylated and there is no indication that it is unstable or that glycosylation increases its stability. Despite their

similarities, Ag43 and AIDA-I might have subtle structural differences that make glycosylation more important for AIDA-I, and it should be noted that the latter seems to be more glycosylated than Ag43. Second, it is surprising that glycosylation is dispensable for auto-aggregation. One possible explanation is that the domains of the protein involved in auto-aggregation do not require the folding and/or the stability provided by glycosylation. The unglycosylated protein could also be degraded into an intermediate that still bears an auto-aggregation domain.

Many uncertainties remain about the glycosylation of AIDA-I. The mechanism of glycosylation itself needs to be investigated in order to understand the specificity of Aah towards its substrates, and more precisely towards specific sites in the latter. How glycosylation affects the conformation of AIDA-I and why glycosylation is not necessary for auto-aggregation is also unclear. Further characterization of the glycosylation of AIDA-I, as well as that of other SAAT, is warranted to tackle these puzzling questions.

#### ACKNOWLEDGMENTS

This work was supported by financial contributions from the Groupe de Recherche et d'Etudes sur les Maladies Infectieuses du Porc (GREMIP), the Natural Sciences and Engineering Research Council of Canada (NSERC discovery grant 262746), the Canada Research Chair program and the Canada Foundation for Innovation (project 201414). M.-E. C. is supported by a graduate fellowship from the Fonds de Recherche sur la Nature et les Technologies du Québec (FQRNT #114663).

#### REFERENCES

1. **Benz, I., and M. A. Schmidt.** 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infect Immun* **57**:1506-11.
2. **Benz, I., and M. A. Schmidt.** 2001. Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Mol Microbiol* **40**:1403-13.
3. **Charbonneau, M. E., F. Berthiaume, and M. Mourez.** 2006. Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I). *J Bacteriol* **188**:8504-12.

4. **Dobos, K. M., K. H. Khoo, K. M. Swiderek, P. J. Brennan, and J. T. Belisle.** 1996. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *J Bacteriol* **178**:2498-506.
5. **Fares, F.** 2006. The role of O-linked and N-linked oligosaccharides on the structure-function of glycoprotein hormones: development of agonists and antagonists. *Biochim Biophys Acta* **1760**:560-7.
6. **Gerwig, G. J., J. P. Kamerling, J. F. Vliegthart, E. Morag, R. Lamed, and E. A. Bayer.** 1993. The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of *Clostridium thermocellum* and *Bacteroides cellulosolvens*. *J Biol Chem* **268**:26956-60.
7. **Grass, S., A. Z. Buscher, W. E. Swords, M. A. Apicella, S. J. Barenkamp, N. Ozchlewski, and J. W. St Geme, 3rd.** 2003. The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol Microbiol* **48**:737-51.
8. **Grizot, S., M. Salem, V. Vongsouthi, L. Durand, F. Moreau, H. Dohi, S. Vincent, S. Escaich, and A. Ducruix.** 2006. Structure of the *Escherichia coli* heptosyltransferase WaaC: binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose. *J Mol Biol* **363**:383-94.
9. **Guerry, P., C. P. Ewing, M. Schirm, M. Lorenzo, J. Kelly, D. Pattarini, G. Majam, P. Thibault, and S. Logan.** 2006. Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. *Mol Microbiol* **60**:299-311.
10. **Ha, S. K., C. Choi, and C. Chae.** 2003. Prevalence of a gene encoding adhesin involved in diffuse adherence among *Escherichia coli* isolates in pigs with postweaning diarrhea or edema disease. *J Vet Diagn Invest* **15**:378-81.
11. **Ha, S. K., C. Choi, K. Jung, J. Kim, D. U. Han, Y. Ha, S. D. Lee, S. H. Kim, and C. Chae.** 2004. Genotypic prevalence of the adhesin involved in diffuse adherence in *Escherichia coli* isolates in pre-weaned pigs with diarrhoea in Korea. *J Vet Med B Infect Dis Vet Public Health* **51**:166-8.
12. **Hellman, U., C. Wernstedt, J. Gonez, and C. H. Heldin.** 1995. Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal Biochem* **224**:451-5.

13. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. C. Fernandez, and D. Ala'Aldeen.** 2004. Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* **68**:692-744.
14. **Herrmann, J. L., P. O'Gaora, A. Gallagher, J. E. Thole, and D. B. Young.** 1996. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. *Embo J* **15**:3547-54.
15. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. *J Biol Chem* **256**:7990-7.
16. **Imperiali, B., and S. E. O'Connor.** 1999. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* **3**:643-9.
17. **Klemm, P., R. M. Vejborg, and O. Sherlock.** 2006. Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol* **296**:187-95.
18. **Kuo, C., N. Takahashi, A. F. Swanson, Y. Ozeki, and S. Hakomori.** 1996. An N-linked high-mannose type oligosaccharide, expressed at the major outer membrane protein of *Chlamydia trachomatis*, mediates attachment and infectivity of the microorganism to HeLa cells. *J Clin Invest* **98**:2813-8.
19. **Laarmann, S., and M. A. Schmidt.** 2003. The *Escherichia coli* AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor. *Microbiology* **149**:1871-82.
20. **Lindenthal, C., and E. A. Elsinghorst.** 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect Immun* **67**:4084-91.
21. **Logan, S. M., J. F. Kelly, P. Thibault, C. P. Ewing, and P. Guerry.** 2002. Structural heterogeneity of carbohydrate modifications affects serospecificity of *Campylobacter* flagellins. *Mol Microbiol* **46**:587-97.
22. **Mainil, J. G., E. Jacquemin, P. Pohl, A. Kaeckenbeeck, and I. Benz.** 2002. DNA sequences coding for the F18 fimbriae and AIDA adhesin are localised on the same plasmid in *Escherichia coli* isolates from piglets. *Vet Microbiol* **86**:303-11.
23. **Mogensen, J. E., and D. E. Otzen.** 2005. Interactions between folding factors and bacterial outer membrane proteins. *Mol Microbiol* **57**:326-46.
24. **Moormann, C., I. Benz, and M. A. Schmidt.** 2002. Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infect Immun* **70**:2264-70.

25. **Mourez, M., S. Skouloubris, J. M. Betton, and E. Dassa.** 1997. Heat shock induction by a misassembled cytoplasmic membrane protein complex in *Escherichia coli*. *Mol Microbiol* **26**:821-31.
26. **Ngeleka, M., J. Pritchard, G. Appleyard, D. M. Middleton, and J. M. Fairbrother.** 2003. Isolation and association of *Escherichia coli* AIDA-I/STb, rather than EAST1 pathotype, with diarrhea in piglets and antibiotic sensitivity of isolates. *J Vet Diagn Invest* **15**:242-52.
27. **Niewerth, U., A. Frey, T. Voss, C. Le Bouguenec, G. Baljer, S. Franke, and M. A. Schmidt.** 2001. The AIDA autotransporter system is associated with F18 and stx2e in *Escherichia coli* isolates from pigs diagnosed with edema disease and postweaning diarrhea. *Clin Diagn Lab Immunol* **8**:143-9.
28. **Oliver, D. C., G. Huang, E. Nodel, S. Pleasance, and R. C. Fernandez.** 2003. A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain. *Mol Microbiol* **47**:1367-83.
29. **Peter-Katalinic, J.** 2005. Methods in enzymology: O-glycosylation of proteins. *Methods Enzymol* **405**:139-71.
30. **Raina, S., D. Missiakas, and C. Georgopoulos.** 1995. The rpoE gene encoding the sigma E (sigma 24) heat shock sigma factor of *Escherichia coli*. *Embo J* **14**:1043-55.
31. **Romain, F., C. Horn, P. Pescher, A. Namane, M. Riviere, G. Puzo, O. Barzu, and G. Marchal.** 1999. Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses. *Infect Immun* **67**:5567-72.
32. **Rutherford, N., M. E. Charbonneau, F. Berthiaume, J. M. Betton, and M. Mourez.** 2006. The periplasmic folding of a cysteineless autotransporter passenger domain interferes with its outer membrane translocation. *J Bacteriol* **188**:4111-6.
33. **Schirm, M., E. C. Soo, A. J. Aubry, J. Austin, P. Thibault, and S. M. Logan.** 2003. Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol Microbiol* **48**:1579-92.
34. **Schmidt, M. A., L. W. Riley, and I. Benz.** 2003. Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol* **11**:554-61.
35. **Sherlock, O., U. Dobrindt, J. B. Jensen, R. Munk Vejborg, and P. Klemm.** 2006. Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. *J Bacteriol* **188**:1798-807.



36. **Sherlock, O., M. A. Schembri, A. Reisner, and P. Klemm.** 2004. Novel Roles for the AIDA Adhesin from Diarrheagenic *Escherichia coli*: Cell Aggregation and Biofilm Formation. *J Bacteriol* **186**:8058-65.
37. **Suhr, M., I. Benz, and M. A. Schmidt.** 1996. Processing of the AIDA-I precursor: removal of AIDAc and evidence for the outer membrane anchoring as a beta-barrel structure. *Mol Microbiol* **22**:31-42.
38. **Szymanski, C. M., and B. W. Wren.** 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* **3**:225-37.
39. **Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust, and P. Guerry.** 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol* **32**:1022-30.
40. **Van den Steen, P., P. M. Rudd, R. A. Dwek, and G. Opdenakker.** 1998. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* **33**:151-208.
41. **Young, N. M., J. R. Brisson, J. Kelly, D. C. Watson, L. Tessier, P. H. Lanthier, H. C. Jarrell, N. Cadotte, F. St Michael, E. Aberg, and C. M. Szymanski.** 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J Biol Chem* **277**:42530-9.

## FIGURE LEGENDS

### **Fig. 1: Localization of heptose modifications.**

Glycosylated peptides identified in this study are indicated on a schematic representation of the pre-protein showing the N-terminal signal sequence required for inner membrane translocation, a region encompassing 35 imperfect 19-amino acids repeats (gray boxes) and the autoproteolytic cleavage site separating mature AIDA-I from the membrane-embedded AIDAc. The imperfectly repeated sequence is highlighted. Numbering corresponds to the position of the amino acids in the pre-protein. In several peptides, a heptose residue could be identified on threonine or serine residues. These residues are circled in the sequence.

### **Fig. 2: Effect of glycosylation on the function of AIDA-I.**

*A.* Adhesion assay. Bacteria bearing an empty vector (-), or expressing unglycosylated AIDA-I (Aida) or glycosylated AIDA-I (Aida/Aah) were inoculated onto monolayer of

confluent Hep-2 cells and adhering bacteria were plated, counted. The *E. coli* strain 2787, from which AIDA-I was originally identified (1), was used as a control. The adhesion is calculated by dividing the cfu of adhering bacteria recovered by the cfu found in the inoculum after 3 hours of incubation. *B.* Biofilm formation. The same strains were grown in minimal medium for 24 hours at 30°C in microtiter plates and biofilms were stained with crystal violet. *C.* Auto-aggregation assay. Cultures of the same strains were left standing at 4°C and the turbidity at the top of the culture was measured. Pictures of the culture tubes were taken after an overnight incubation at 4°C. Experiments were performed at least three times in duplicate (*A, C*) or quadruplicate (*B*) and the values represent means  $\pm$  SEM.

**Fig. 3: Effect of glycosylation on the expression level and on the mild heat extraction of AIDA-I.**

*A.* Whole cell extracts of bacteria bearing an empty vector (-), or expressing unglycosylated AIDA-I (Aida) or glycosylated AIDA-I (Aida/Aah) were separated by SDS-PAGE and probed by immunoblotting with antiserum against the HisG tag fused to mature AIDA-I. *B.* Bacteria harboring the same plasmids were heat-extracted and the extracts recovered by centrifugation prior to separation by SDS-PAGE and Coomassie Blue staining.

**Fig. 4: Effect of glycosylation on the resistance to proteases.**

Bacteria bearing an empty vector (-), or expressing unglycosylated AIDA-I (Aida) or glycosylated AIDA-I (Aida/Aah) were pelleted and resuspended in PBS in the presence or absence of trypsin or proteinase K (0.3  $\mu\text{g}\cdot\text{ml}^{-1}$  and 3  $\text{mg}\cdot\text{ml}^{-1}$ ). The amount of AIDA-I was revealed by immunoblotting with an anti-HisG antibody (upper panel), and the intensity of the bands was quantitated by densitometry and normalized (lower panel).

**Fig. 5: Induction of an extracytoplasmic stress by the non-glycosylated AIDA-I.**

An empty vector (-), or plasmids allowing expression of unglycosylated AIDA-I (Aida) or glycosylated AIDA-I (Aida/Aah) were transformed into the SR1364 (*A*) and SR1458 (*B*) reporter strains. The  $\beta$ -galactosidase activity measured in these strains and presented in Miller units indicate cytoplasmic and extracytoplasmic stress, respectively. The experiment was performed three times in duplicate. The values represent means  $\pm$  SEM and were compared by an ANOVA (\*,  $p < 0.01$ ).

**Fig. 6: Binding of unglycosylated mature AIDA-I to cultured epithelial cells.**

The glycosylated (open squares) or unglycosylated (closed squares) fusion of GST to mature AIDA-I (encompassing amino acids 50 to 847) were purified by affinity chromatography. Various amounts of the proteins were incubated with fixed Hep-2 cells. Pure GST (diamonds) and pure whole AIDA-I (crosses) were used as a control. Bound proteins were revealed with an anti-GST antiserum directly coupled to HRP or an anti-AIDA-I serum and a secondary antibody coupled to HRP. Background was subtracted and the absorption was normalized by dividing the absorption of each well by the maximal absorption measured on the plate in order to obtain percentages of maximal binding.

Peptide		Number of heptoses	Means of identification
Glycosylated peptides			
P1	K <sup>τ</sup> T <sub>95</sub> TATTVNSSGSQNVGTSGATISTIVNSSGGIQR <sub>126</sub> <sup>τ</sup>	2,3 or 4	MS, ED
P2	Y <sup>τ</sup> N <sub>147</sub> LGHASNTIVIF <sub>157</sub> <sup>τ</sup>	0 or 1	MS
P3	R <sup>τ</sup> V <sub>240</sub> N <sub>2</sub> SGAVATGTVL <sub>2</sub> SGG-	ND	ED
P4a	K <sup>τ</sup> G <sub>328</sub> SQIVN <sub>2</sub> SEGTAINTLVSDGGYQHIR <sub>352</sub> <sup>τ</sup>	0,1 or 2	MS, ED
P4b	L <sup>τ</sup> S <sub>323</sub> ANIKGSQIVN <sub>2</sub> SEGTAINTLVSDGGY <sub>348</sub> <sup>τ</sup>	2	MS
P5	R <sup>τ</sup> V <sub>389</sub> L <sub>2</sub> SDGYAR <sub>396</sub> <sup>τ</sup>	ND	ED
P6	R <sup>τ</sup> E <sub>406</sub> NV <sub>2</sub> N <sub>2</sub> GGVSYNAM-	ND	ED
P7	Y <sup>τ</sup> I <sub>427</sub> YSDGEATAAIVNTSGF <sub>443</sub> <sup>τ</sup>	1	MS
P8a	R <sup>τ</sup> Q <sub>536</sub> YVY <sub>2</sub> SGATATSTVGNNEGR <sub>554</sub> <sup>τ</sup>	0, 1 or 2	MS, ED
P8b	Y <sup>τ</sup> V <sub>538</sub> YSGATATSTVGNNEGREGY <sub>556</sub> <sup>τ</sup>	1,2	MS
P9	R <sup>τ</sup> E <sub>555</sub> YVL <sub>2</sub> SGGITDGTVLN <sub>2</sub> SGGLQAVSSGGK <sub>581</sub> <sup>τ</sup>	4	MS, ED
P10	K <sup>τ</sup> A <sub>582</sub> SATVINEGGAQFVYDGGQVTGTNIK <sub>607</sub> <sup>τ</sup>	0 or 1	MS, ED
Unglycosylated peptides			
	Y <sup>τ</sup> Q <sub>349</sub> HIRNNGGIASGTIVN <sub>2</sub> QSGY <sub>367</sub> <sup>τ</sup>	-	MS
	R <sup>τ</sup> G <sub>397</sub> TILNNSGR <sub>405</sub> <sup>τ</sup>	-	ED
	K <sup>τ</sup> A <sub>470</sub> IDAEVYSSGGK <sub>480</sub> <sup>τ</sup>	-	ED
	Y <sup>τ</sup> S <sub>485</sub> GGEVSGTQIF <sub>495</sub> <sup>τ</sup>	-	MS
	R <sup>τ</sup> L <sub>518</sub> NAFAGNVVGTILNQEGR <sub>535</sub> <sup>τ</sup>	-	MS, ED
	K <sup>τ</sup> D <sub>739</sub> NTGIMTYAGTLTQAQGVNK <sub>759</sub> <sup>τ</sup>	-	MS, ED
	K <sup>τ</sup> L <sub>824</sub> LLSATVNGSLVNNK <sub>837</sub> <sup>τ</sup>	-	MS, ED
	K <sup>τ</sup> N <sub>838</sub> NIILNPTK <sub>845</sub> <sup>τ</sup>	-	ED

**Table I: Glycosylation of peptides identified in AIDA-I.**

Glycosylated mature AIDA-I (corresponding to amino acids 50 to 847 in the pre-protein) was processed by digestion with trypsin or chymotrypsin and the resulting peptides identified by MS or ED. In MS, glycosylation of a peptide was identified as an

excess of mass corresponding to a multiple of 192 Da (the mass of one heptose residue). In ED, glycosylated residues are identified when a signal corresponding to the modified residue is lacking or reduced during a degradation cycle. Two residues (S577, S578) showed only an 80% reduction in signal, suggesting that these residues were not modified in all peptides. The sequences of peptides P1, P3 and P6 could not be completely obtained by ED. Numbering corresponds to the position of the amino acids in the pre-proprotein. MS yielded the number of heptose molecules bound per peptide. However, except in one case (T154), MS did not permit to identify the modified residues. Using ED, the number of heptoses cannot be determined (not determined, ND) but several residues (S102, S111, S116, S242, S252, S334, S391, S409, S540, S546, S559, S570, S577, S578 and S583) were identified as modified.

Figure 1

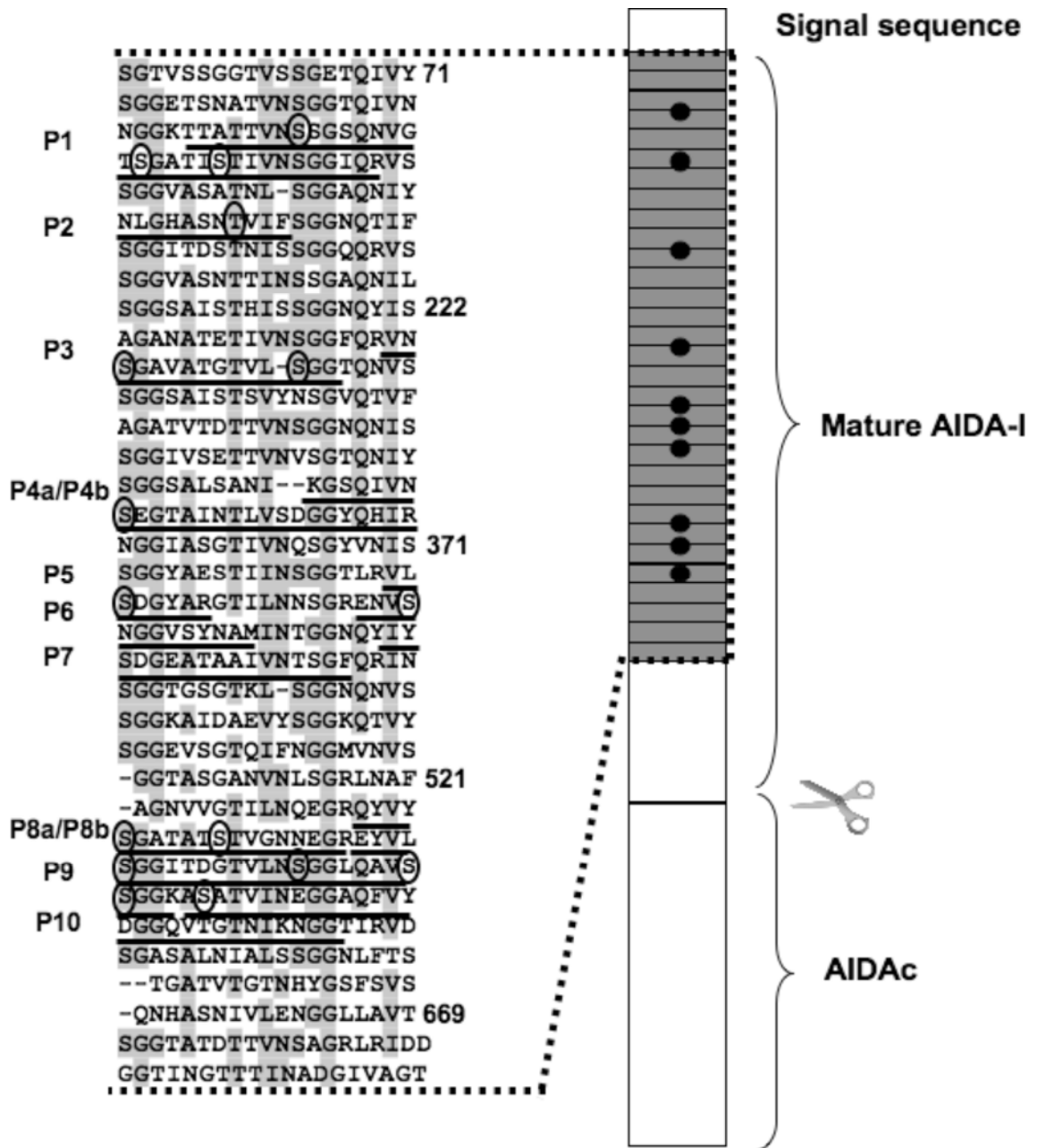


Figure 2

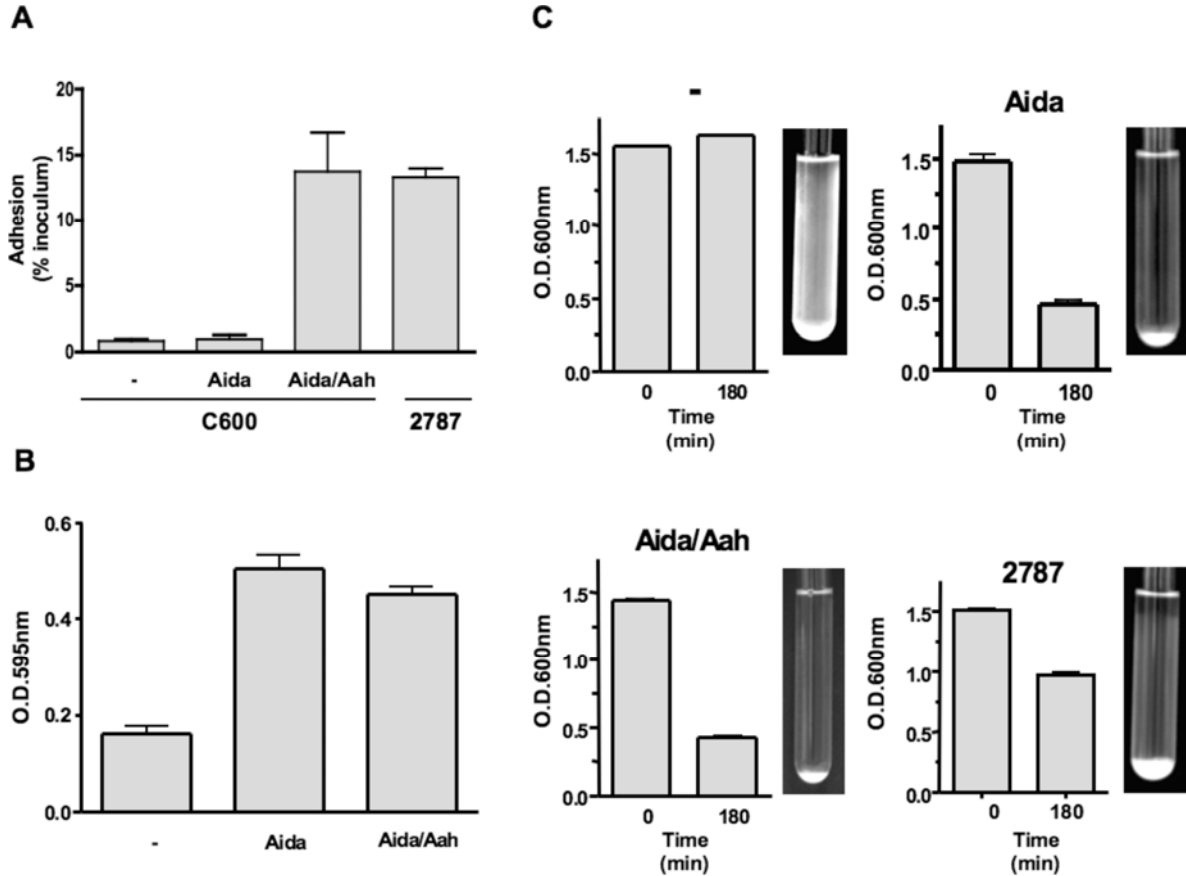


Figure 3

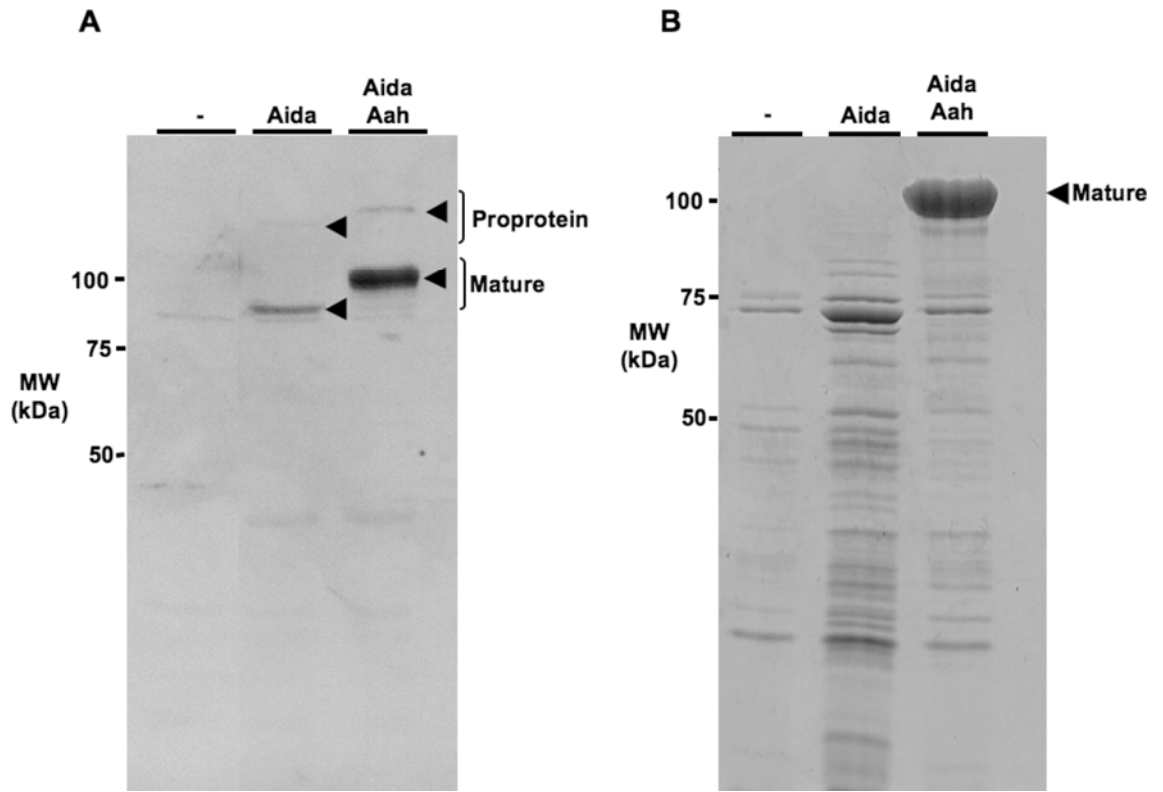




Figure 4

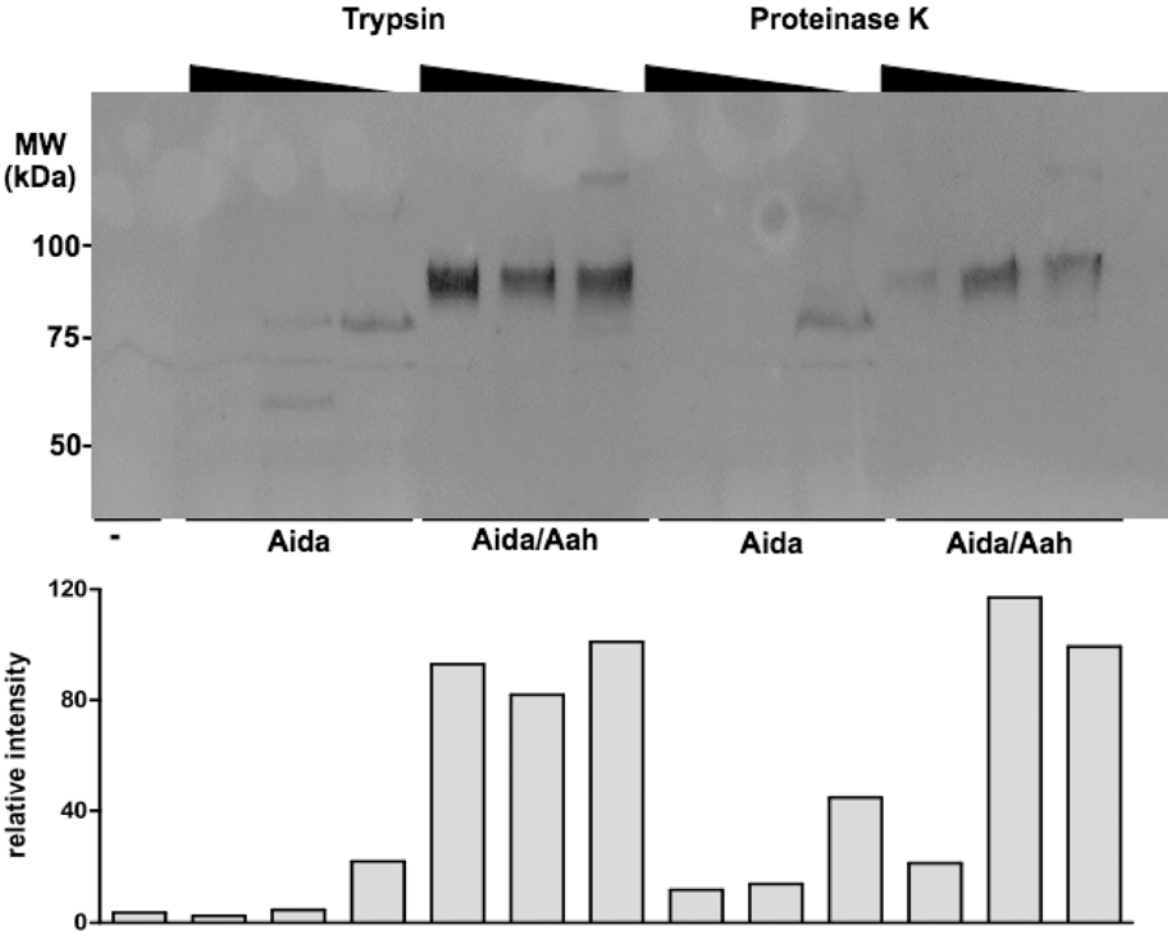


Figure 5

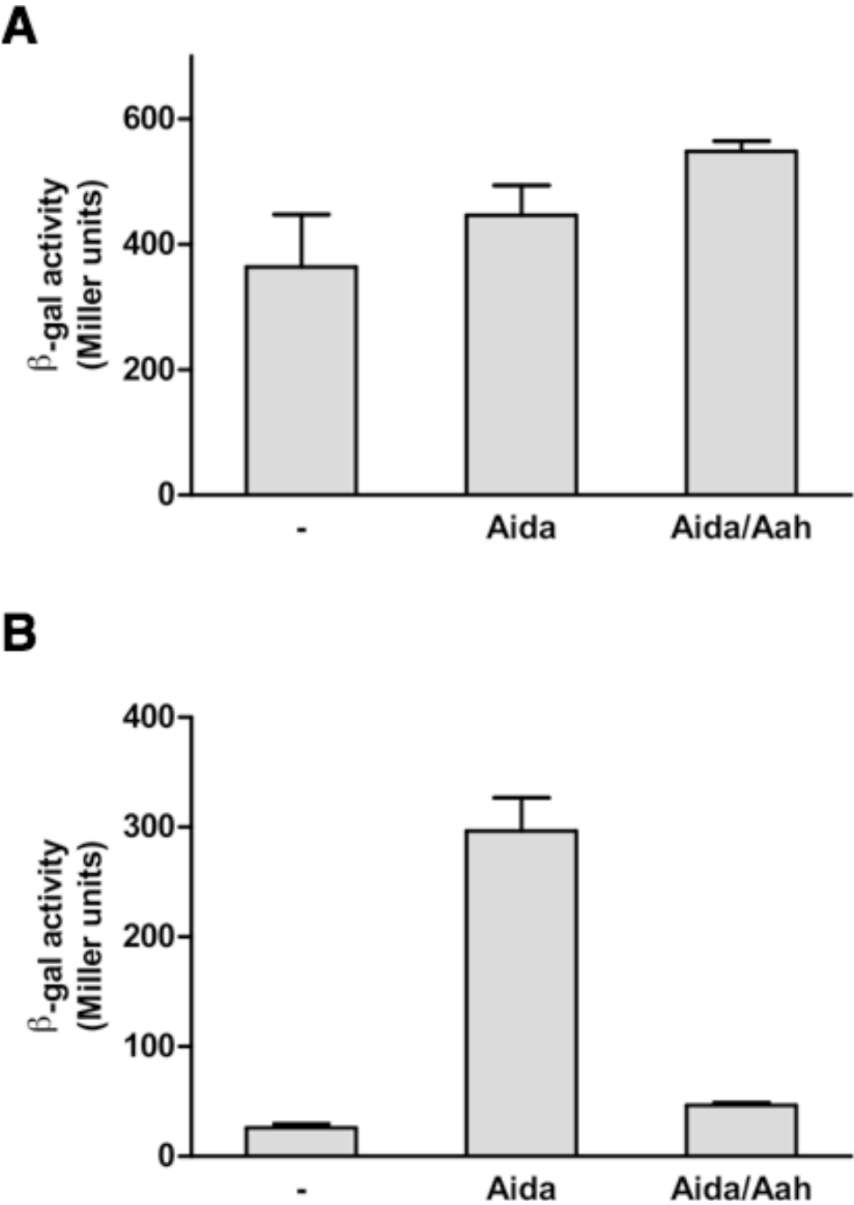
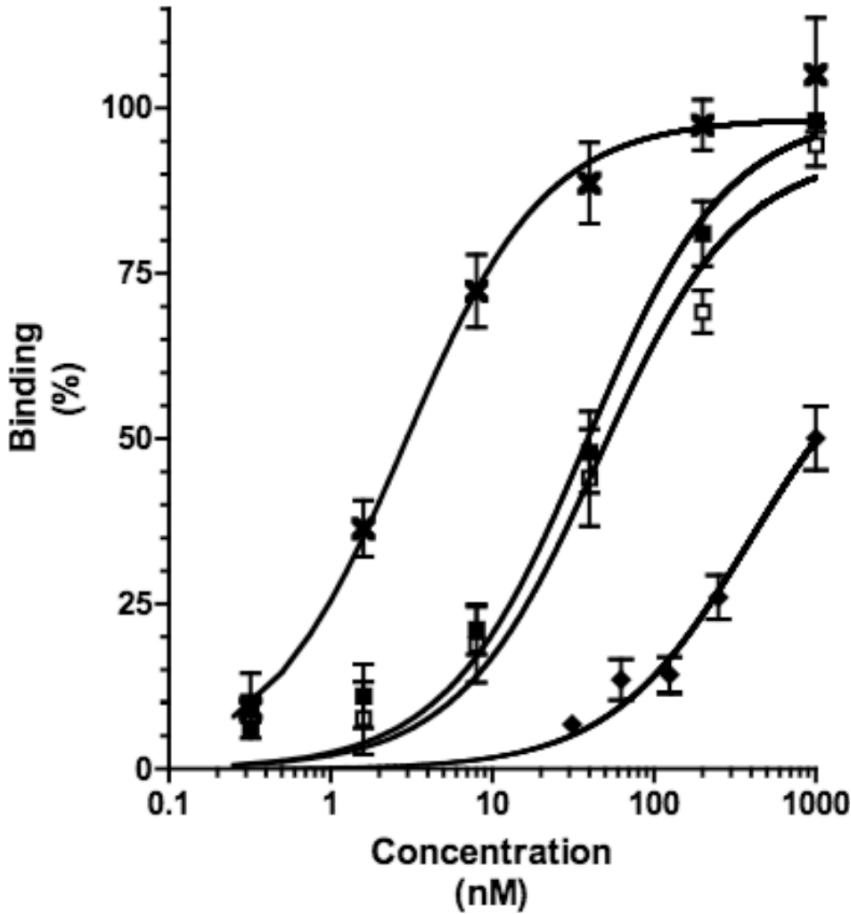
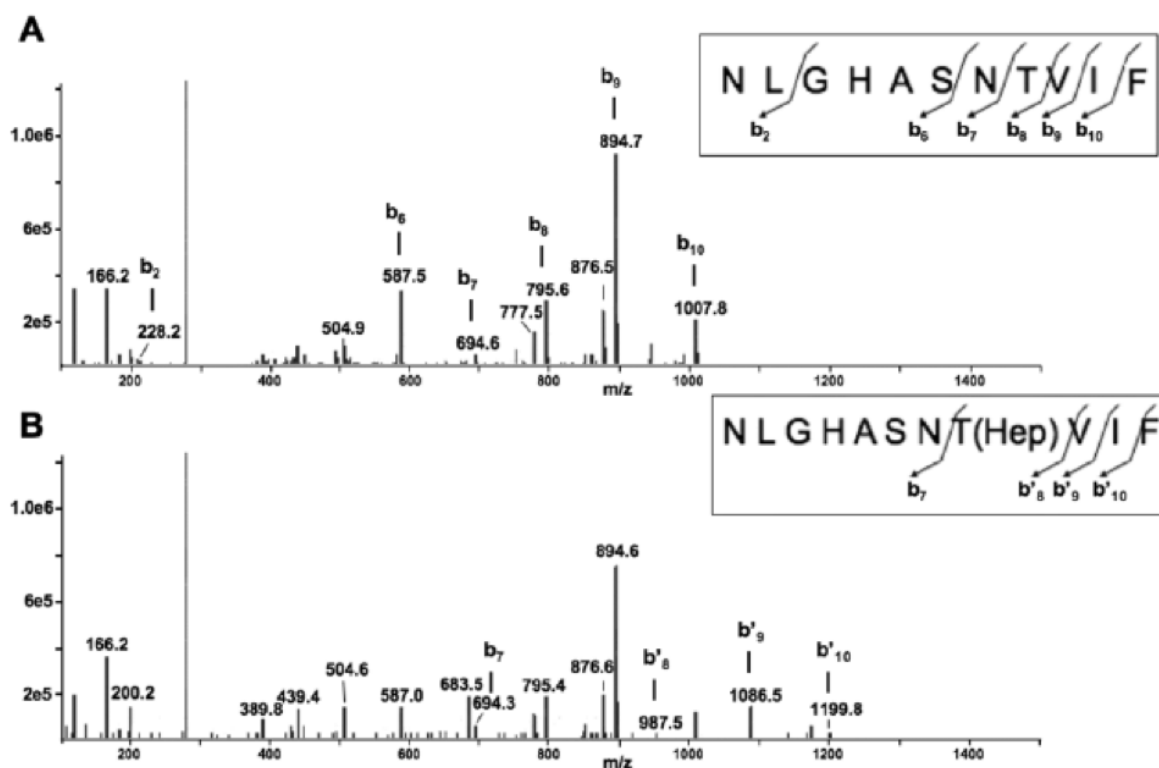


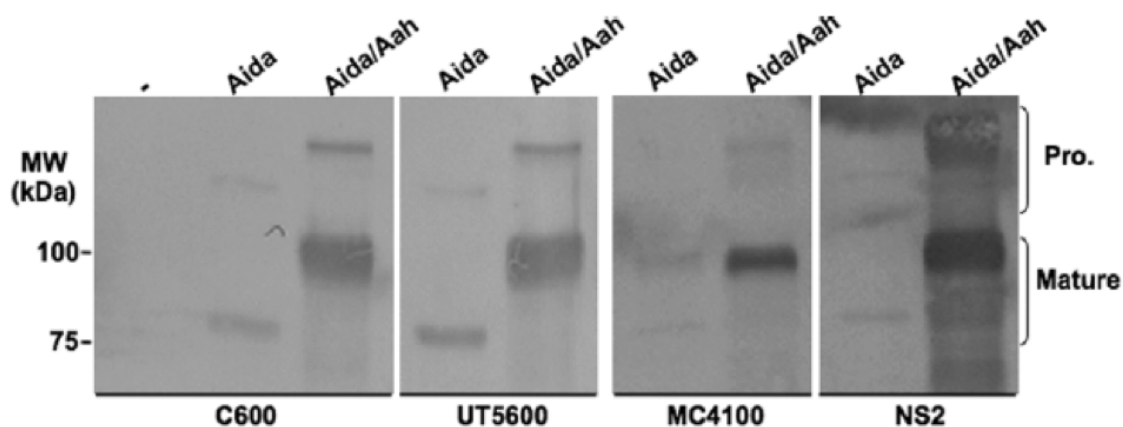
Figure 6



## Supporting information



**FIG. S1: Modification of a threonine residue in a peptide obtained by treatment of AIDA-I with chymotrypsin.** A. CID spectrum of the doubly charged  $m/z$  587 ion (molecular weight 1172 Da) corresponding to the sequence displayed. The highlighted fragment ions correspond to the C terminal part of the peptides with sequential cleavage of the peptide backbone with the residual charge residing on the N-terminal end of the peptide (b-ions). B. CID spectrum of the doubly charged  $m/z$  683 ion (molecular weight 1364 Da) corresponding to the same sequence but for the presence of a heptose residue on the threonine. The same b ions at the C terminal part of the peptides are observed, but with the addition of new  $b_8'$ ,  $b_9'$  and  $b_{10}'$  ions which show a mass increment of 192 Da respective to A, proving that the heptose residue is located on the threonine.



**Fig. S2: DegP and OmpT are not involved in the degradation of unglycosylated AIDA-I.**

In order to assess the effect of the extracytoplasmic proteases DegP and OmpT on the levels of production of glycosylated and unglycosylated AIDA-I, we compared the amounts of these proteins in the *Escherichia coli* K12 strains C600 (F<sup>-</sup>, thr-1, leuB6, thi-1, lacY1, supE44, rfbD1, fhuA21), UT5600 (F<sup>-</sup>, ara-14, leuB6, secA6, lacY1, proC14, tsx-67,  $\Delta$ (ompT-fepC)266, entA403, trpE38, rfbD1, rpsL109, xyl-5, mtl-1, thi-1), MC4100 (F<sup>-</sup>, araD139,  $\Delta$ (argF-lac)U169, rpsL150, relA1, flbB5301, deoC1, ptsF25, rbsR) and NS2 (MC4100  $\Delta$ malE444, degP41, recA1, srl::Tn10, a generous gift from Dr Jean-Michel Betton, Institut Pasteur, Paris). This panel of strains compares strains that can express DegP and OmpT (C600, MC4100) with a strain that can only express DegP (UT5600) and a strain that can only express OmpT (NS2). Overnight cultures from these strains harboring pTRC99A, pAngH (Aida, non-glycosylated AIDA-I) or pAngH and pAah (Aida/Aah, glycosylated AIDA-I) were grown at 30°C, and induced overnight with 10  $\mu$ M of IPTG when they reached an O.D.<sub>600nm</sub> of 0.8. The bacterial cultures were normalized to the same O.D.<sub>600nm</sub> and equal amounts of cultures were centrifuged for 10 minutes at 12,000 x g in microfuge tubes. The pellets were resuspended in PBS and diluted in twice concentrated SDS-PAGE loading buffer containing  $\beta$ -mercaptoethanol immediately before heating at 100°C for 10 minutes. The samples were then separated by SDS-PAGE on 10% acrylamide gels and the gels were transferred to polyvinylidene fluoride membranes. Immunodetection was performed with an anti-HisG HRP-coupled antibody. Immune complexes were revealed using a 3,3',5,5'-tetramethylbenzidine solution for membranes and showed mature AIDA-I and the proprotein (Pro.).

From this experiment it is apparent that the expression of unglycosylated or glycosylated AIDA-I is not improved in the absence of OmpT (compare MC4100 or C600 with UT5600). Higher amounts of glycosylated protein are found in the absence of DegP (compare MC4100 and NS2). The expression of unglycosylated protein is also slightly increased but remains barely detectable.

Taken together, these results suggest that neither DegP nor OmpT are responsible for the degradation of unglycosylated AIDA-I.

## ARTICLE 4

**The *Escherichia coli* AIDA-I autotransporter undergoes glycosylation independently of export**

Research in Microbiology, 2008, Sept.-Oct., Volume 159 (7-8), pages 537-44

Contribution de l'auteur pour cet article

En tant que premier auteur de cet article, j'ai effectué toutes les expérimentations décrites et j'ai écrit la totalité de l'article.

Contribution scientifique de l'article

Cet article correspond à la première caractérisation d'une glycosyltransférase ayant comme cible des protéines chez *E. coli*. Nous avons démontré que Aah était une enzyme cytoplasmique qui permet la glycosylation de AIDA-I indépendamment de son transport à travers la membrane interne. Une glycosylation cytoplasmique est rare chez les bactéries Gram négatif, avec un seul exemple existant dans la littérature au moment de la publication.

**The *Escherichia coli* AIDA-I autotransporter undergoes cytoplasmic glycosylation independently of export**

Marie-Ève Charbonneau and Michael Mourez\*

Canada Research Chair on Bacterial Animal Diseases, Université de Montréal, Faculté de Médecine Vétérinaire, 3200 Sicotte, St-Hyacinthe, J2S 7C6, Québec, Canada

Keywords: AIDA-I; autotransporter; signal sequence; glycosylation; *Escherichia coli*



**Abstract**

The *Escherichia coli* Adhesin Involved in Diffuse Adherence (AIDA-I) is an outer membrane autotransporter protein and one of the few glycosylated proteins found in *Escherichia coli*. O-glycosylation is mediated by the product of the *aah* gene, which codes for a heptosyltransferase that uses ADP-glycero-manno-heptose precursors from the LPS biosynthesis pathway. Little else is still known about Aah and the mechanism involved in the modification of AIDA-I. We observed that Aah is mainly found in an insoluble fraction and, by deletion of AIDA-I signal sequence or by blocking the sec-translocation machinery with sodium azide, we showed that glycosylation occurs in the cytoplasm of bacteria independently of secretion. Since AIDA-I harbors a N-terminal extension in its signal sequence, we wondered if glycosylation requires this unusual sequence. We observed that while deletion of the N-terminal extension affected the expression level of AIDA-I, the protein was still exported to the outer membrane and glycosylated. However, modification of a secreted protein in the cytoplasm raises several mechanistic questions.

## Introduction

The adhesin involved in diffuse adherence (AIDA-I) of *Escherichia coli* is an afimbrial adhesin that mediates diffuse adherence to epithelial cells. It was originally cloned from a strain isolated in a case of infantile diarrhea [3]. Since then, AIDA-I has mainly been found in association with pathogenic strains of *E. coli* causing edema disease and post weaning diarrhea in piglets, which represent major causes of economic losses in farms worldwide [6, 33].

AIDA-I belongs to the monomeric autotransporter protein family of the type V secretion pathway, which includes an important number of proven and suspected virulence factors [15, 16]. A common feature of autotransporters is their modular organization. AIDA-I is synthesized from the *aidA* gene as a 145 kDa pre-proprotein with three domains (Fig. 1) [4]. First, a N-terminal *sec*-dependent signal sequence that allows transport across the inner membrane. Second, a central domain of 100 kDa, the mature AIDA-I, bearing the functional part of the protein. Lastly, a C-terminal membrane-embedded domain of 45 kDa, AIDA<sub>c</sub>, which is predicted to form a pore in the outer membrane and might serve as a translocation conduit for mature AIDA-I. Mature AIDA-I is cleaved from the proprotein but remains strongly associated with AIDA<sub>c</sub> at the bacterial surface [7].

The *aidA* gene is associated with a second gene coding for a glycosyltransferase and named *aah*, for autotransporter adhesin heptosyltransferase [5]. Two other autotransporters homologous to AIDA-I, the Ag43 aggregation factor and the TibA invasin, can also be glycosylated [22, 23, 27]. Glycosylation can be performed by Aah or TibC, another glycosyltransferase homologous to Aah and associated with TibA. Aah and TibC do not appear to modify any other protein and AIDA-I, Ag43 and TibA have been proposed to form a subfamily of functionally and structurally related autotransporters called SAATs for Self-Associating AutoTransporters [20]. The Aah glycosyltransferase uses ADP-glycero-manno-heptose precursors from the LPS biosynthesis pathway [5]. Heptose molecules can be found on at least 16 serine or threonine residues of mature AIDA-I and the glycosylation seems to be highly heterogeneous [8]. Eleven peptides in the extracellular part of Ag43 can also be heterogeneously glycosylated by Aah [27]. However, the mechanism of action of these glycosyltransferases is unknown.

Most bacterial glycoproteins are membrane or secreted proteins and glycosylation usually occurs after the crossing of the inner membrane [30]. This is similar to the situation in

eukaryotes, where glycosylation takes place after crossing the endoplasmic reticulum membrane [9]. Consequently, most protein glycosyltransferases face an extracytoplasmic compartment. Where AIDA-I, Ag43 or TibA glycosylation occurs, however, is not known and the cellular localization of Aah or TibC has not been determined.

Like several other proteins belonging to the type V secretion pathway, the AIDA-I, Ag43 and TibA proteins possess an unusually long signal sequence that often exceeds 50 amino acids [11, 17]. This atypical length is the result of a N-terminal extension (Fig. 1), the sequence of which is conserved [11]. Several studies have been performed to determine the function of this unusual extension, but it remains controversial [10, 12, 18, 25]. In the case of AIDA-I, Ag43 and TibA, the extension could be important for coupling translocation across the inner membrane with glycosylation. Such a coupling has been extensively described for the eukaryotic oligosaccharyl transferases that mediate N-glycosylation of polypeptide during translocation into the endoplasmic reticulum [9]. Several instances also exist in bacteria: the coupling of proteins O-mannosylation to Sec-dependent secretion has been demonstrated in *Mycobacterium tuberculosis* [31], and coupling of secretion with glycosylation has been suggested in the biogenesis of the Fap1 adhesin of *Streptococcus parasanguis* [32]. Interestingly, the secretion and glycosylation of the GspB adhesin of *Streptococcus gordonii* is also coordinated and involves an unusually long signal sequence [2].

In this study, we wanted to determine the subcellular location where glycosylation of AIDA-I occurs and assess the possible role of the N-terminal extension of the signal sequence. To answer these questions, we constructed a deletion of the whole AIDA-I signal sequence or a deletion of its N-terminal extension, and we assessed the biogenesis and glycosylation of the mutant proteins. We observed that AIDA-I is glycosylated in the cytoplasm. The presence of the N-terminal extension influenced the levels of protein expression, but the extension was dispensable for the glycosylation process itself.

## **Material and methods**

### **Bacterial strain and plasmids**

The *Escherichia coli* K12 strains C600 (F<sup>-</sup>, *thr1*, *leuB6*, *thi1*, *lacY1*, *supE44*, *rfbD1*, *fhuA21*), UT5600 (F<sup>-</sup>, *ara14*, *leuB6*, *secA6*, *lacY1*, *proC14*, *tsx67*,  $\Delta(\textit{ompT-fepC})266$ , *entA403*, *trpE38*, *rfbD1*, *rpsL109*, *xyl5*, *mtl1*, *thi1*) and BL21 (F<sup>-</sup>, *ompT*, *hsdS*(<sub>T<sub>B</sub></sub><sup>-</sup>m<sub>B</sub><sup>-</sup>), *gal*) were used in this study. Plasmids pAngH and pAah allowing the expression, respectively,

of AIDA-I and Aah, have been previously described [8]. The cloning procedure for pAngH results in a modification of the amino acid in position 2, replacing the asparagine by a glycine, and introduces a protein tag consisting of six histidine residues and a glycine (HisG) at the N terminus of the proprotein. Plasmid pAngH is derived from the pTRC99A vector (Pharmacia Biotech) and is compatible with pAah, a pACYC184-derived plasmid. In all constructs, protein expression is under the control of isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) through the *ptrc* promoter.

For deletion of the whole signal sequence ( $\Delta$ SS) we amplified from plasmid pAngH the portion of the *aida* gene downstream of the signal peptide, using the primers 5'-TATACCATGGTAAATATTTTCAGGCCATCATCATCAT-3' and 5'-TTATCTAGATTATCAGAAGCTGTATTTTATCC-3', which introduced recognition sites for restriction enzymes *Nco*I and *Xba*I. The resulting fragment was cloned in the pTRC99a vector (Pharmacia Biotech), generating pAngH $\Delta$ SS (Fig. 1). To introduce a deletion of the N-terminal extension (SS $\Delta$ N), we used the Quick-Change II site-directed mutagenesis kit (Statagene) on plasmid pAngH, using the primers 5'-GGAAACAGACCATGGGTAGAGGACATGGTTTTGTCC-3' and its corresponding complementary oligonucleotide, thus generating pAngHSS $\Delta$ N (Fig. 1). Lastly, to generate the pAahH plasmid, a six-histidines tag was introduced at the C-terminal end of Aah by site-directed mutagenesis of the pAah plasmid using the primers 5'-CATGACAGGATCATTCATCATCATCATCATTAAGGATCCTCTAG-3' and its complementary oligonucleotide.

### **Bacterial growth and induction conditions**

Bacteria containing the different plasmids were grown at 30°C on LB agar plates or in liquid LB medium containing 100  $\mu$ g.ml<sup>-1</sup> ampicillin and 50  $\mu$ g.ml<sup>-1</sup> chloramphenicol. Growth was followed by measuring turbidity as the optical density at 600 nm (OD<sub>600nm</sub>). At an OD<sub>600nm</sub> of 0.8, the cultures were induced with 10  $\mu$ M IPTG, unless indicated otherwise. This low concentration of IPTG was used to limit the toxicity associated with overexpression of AIDA-I and/or Aah. In experiments with sodium azide, bacteria were induced with 50  $\mu$ M of IPTG at an OD<sub>600nm</sub> of 0.4 and were grown in presence of 0.02% sodium azide for 90 minutes.

### **Heat extraction**

Heat extracts were obtained as previously described [7]. Briefly, overnight cultures from *E. coli* C600 bearing appropriate plasmids were normalized at the same OD<sub>600nm</sub> in 10 ml of LB broth. Bacteria were harvested and resuspended in 150 µl of 10 mM sodium phosphate buffer, pH 7. In order to release the mature AIDA-I, the samples were heated at 60°C for 20 minutes. The treated samples were centrifuged for 5 minutes at 12,000 x g in microfuge tubes. A 10 µl aliquot from the supernatant was submitted to SDS-PAGE and stained with Coomassie blue, as described below.

### **Whole cell extracts and cellular fractionation**

For whole cell extracts, overnight cultures were grown, normalized and centrifuged for 10 minutes at 12,000 x g in microfuge tubes and the pellets were resuspended in 100 µl of Phosphate buffered saline (PBS). In all experiments comparing levels of protein expression, samples correspond to identical amounts of bacteria, normalized using OD<sub>600nm</sub>. Cellular fractionation was performed as previously described [26]. Briefly, bacterial cultures were normalized, harvested and resuspended in 950 ml of 10 mM Tris-HCl pH 8, 0.75 mM sucrose, 0.4 mg.ml<sup>-1</sup> lysozyme and 10 mM EDTA pH 8. The resulting spheroplasts were collected, resuspended in Tris-Buffered Saline (TBS, 50 mM Tris-HCl pH 8, 150 mM NaCl) containing proteases inhibitor (Complete mini, Roche) and lysed using an ultrasonic processor. Cellular debris were removed by low speed centrifugation. The lysates were centrifuged for 1 hour in an ultracentrifuge at 250,000 x g. The pellets, containing insoluble fractions, were resuspended in TBS. The volumes of all fractions are normalized in order to correspond to an identical amount of initial bacterial culture. For differential solubilization, the insoluble fractions obtained above were resuspended in TBS containing 1% Triton X-100, incubated for one hour and centrifuged for 1 h in an ultracentrifuge at 250,000 x g. The pellets were then resuspended in 8M urea, incubated for one hour and centrifuged again as before. For sedimentation over a sucrose gradient, the insoluble fractions were resuspended in 10 mM HEPES buffer and a 100-µl aliquot was layered on top of a tube filled with 350 µl of 65% sucrose, 300 µl of 42% sucrose and 250 µl of 24% sucrose. The sample was centrifuged for 4 h at 250,000 x g, and separated in 100-µl fractions.

All samples were processed by SDS-PAGE and immunoblotting, as described below. To control the quality of our fractionation, immunoblotting was also performed to reveal the cytoplasmic protein GroEL and the periplasmic enzyme β-lactamase (data not shown).

**SDS-PAGE, immunoblotting, glycan detection and N-terminal sequencing.**

Protein samples were diluted in twice-concentrated SDS-PAGE loading buffer containing  $\beta$ -mercaptoethanol, and denatured by heating at 100°C for 10 minutes. The samples were then separated by SDS-PAGE on 10% acrylamide gels. The gels were either stained with Coomassie blue or transferred to polyvinylidene fluoride membranes (Millipore). Immunodetection was performed with a serum raised against the glycosylated heat-extracted mature AIDA-I, diluted 1:10,000 in blocking buffer (5% skim milk, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Triton X-100). A goat anti-rabbit horseradish peroxidase (HRP) conjugated antibody (Sigma) was used as a secondary antibody, according to the instructions of the manufacturer. Alternatively, an anti-HisG HRP-coupled antibody (Invitrogen) diluted 1:5,000 in blocking buffer was used for the detection of proteins containing the HisG tag. As controls during subcellular fractionation, antibodies against GroEL protein (Sigma) or against  $\beta$ -lactamase (Chemicon international) were also used. Immune complexes were revealed using a 3,3',5,5'-Tetramethylbenzidine solution for membranes (Sigma). The detection of glycosylation was accomplished using the digoxigenin-hydrazine labeling glycan detection kit (DIG, Roche), according to the instructions of the manufacturer. N-terminal degradation was performed at the Biotechnology Research Institute (Montréal, Québec, Canada).

**RNA isolation and real time PCR**

Strains were grown to stationary phase overnight and then subcultured 1:100, induced with 10  $\mu$ M IPTG after 75 minutes and grown until they reached an OD<sub>600nm</sub> of 0.6. RNA was extracted using the RiboPure-Bacteria kit (Ambion), according to the instructions of the manufacturer. Quantifications were performed using the QuantiTech SYBR green RT-PCR kit (Qiagen) and a Smart Cycler II (Cepheid) with 500 ng of RNA. Thermal cycling conditions consisted of an initial step at 50°C for 30 minutes, an initial PCR activation step at 95°C for 15 minutes, following by 40 cycles of denaturation (94°C, 15 seconds), annealing (55°C, 20 seconds) and extension (72°C, 30 seconds). The sequences of the primers used are: *rpoD*-F: 5'-TCGATACCGAAACGCATACGCAGA-3', *rpoD*-R: 5'-ACGCCGATCGGTGATGATGAAGAT-3', for the *rpoD* control gene, and for the *aidA* gene *aidA*-F: 5'-TCTTCCGGAGGAATTGTGTCAGAGA-3' and *aidA*-R: 5'-GGATGCAATGCCACCGTTTCTGAT-3'. To calculate the fold changes in mRNA amounts, the average cycle threshold (Ct) for the control gene (*rpoD*) was subtracted from

the average Ct for the experimental gene (*aidA*), resulting in the  $\Delta$ Ct.  $\Delta$ Ct obtained in two different backgrounds were then subtracted to obtain the  $\Delta\Delta$ Ct and fold changes are calculated as  $2^{-\Delta\Delta\text{Ct}}$ . The results are the average of three different RNA extraction and quantification experiments.

## Results and discussion

### Subcellular location of the Aah glycosyltransferase

The SignalP 3.0 server (<http://www.cbs.dtu.dk/services/SignalP/>, [1]) fails to identify a signal sequence in the Aah or TibC glycosyltransferases. The PSORTb 2.0 server (<http://www.psort.org/>, [13]) also fails to predict a subcellular location for these proteins, and this question has not been experimentally addressed. As shown in Fig. 2A, when comparing the insoluble fractions of bacteria harboring an empty vector or a plasmid allowing the expression of Aah, an additional protein band with an approximate molecular weight of 40 kDa appears in the strain expressing Aah. The N-terminal sequencing of this protein revealed the sequence MEFTFLSPP, confirming that it is Aah. Only a faint amount of Aah could be seen in the cytoplasm. The localization of Aah in the insoluble pellet is not affected by the presence of AIDA-I (data not shown).

This result suggests that Aah is localized in the inner membrane. Alternatively, this localization could be the result of our expression system, which causes a significant overexpression and, possibly, aggregation. To ascertain the localization of Aah, we performed another fractionation where we sequentially used Triton X-100 and Urea to treat the insoluble fraction, as previously described [24]. We used a construct where a 6-histidines tag was inserted at the C terminus of the protein, in order to reveal the protein through immunoblotting. As show in Fig. 2B, a substantial amount of Aah was found in the cytoplasm. The insoluble material was recovered in almost equal parts in the three fractions recovered during the Triton and Urea treatments: the Triton-soluble fraction, which corresponds to protein associated with the inner membrane, the Triton-insoluble and Urea-soluble fraction, corresponding to aggregated protein, and the Triton- and Urea-insoluble fraction, corresponding to outer membrane proteins. It is unlikely that Aah is really located to the outer membrane, since it does not bear signal sequence. Therefore the Aah recovered in this last fraction is most likely membrane-associated protein that resisted solubilization with Triton X-100 or aggregated protein. The glycosyltransferases that modify LPS, to which Aah is highly homologous, are known to be inner membrane proteins that are poorly

solubilized by Triton X-100 and tend to aggregate [19]. Centrifugation on an isopycnic sucrose density gradient of the insoluble material from whole cell lysates agreed well with those results and showed that about one third of Aah is associated with the inner membrane, and two thirds with particles of higher densities that pelleted at the bottom of the tube (data not shown). Taken together, our results therefore suggest that Aah is found in equilibrium between the cytoplasm and the inner membrane, with some material in found in aggregates, most likely because of the overexpression of the protein.

### **Construction of a mutant of AIDA-I lacking its signal sequence**

To test if AIDA-I could be glycosylated in the cytoplasm, we expressed AIDA-I without its signal sequence in the *E. coli* strain C600 by deleting amino acids Gly2 to Ala49 ( $\Delta$ SS, Fig. 1), and we co-expressed Aah.

We probed whole cell extracts with antibodies against the HisG tag or against mature AIDA-I. As shown in Fig.3A, the  $\Delta$ SS protein is expressed, but at a lower level than the wild-type protein. Quantification of *aida* mRNA by quantitative reverse transcription polymerase chain reaction (qRT-PCR) indicated that the transcript level for the  $\Delta$ SS mutant was 2.2-fold lower than for wild-type AIDA-I. This decrease in transcription is most likely not significant enough to account for the reduction in protein levels observed in Fig. 3A. We expressed the  $\Delta$ SS mutant in the *E. coli* strain BL21, which lacks the cytoplasmic protease Lon and the outer membrane proteases OmpT and OmpP, or in the *E. coli* strain UT5600, which lacks only OmpT and OmpP. As shown in Fig. 3A, the expression levels of the wild-type and  $\Delta$ SS proteins in the BL21 strain were augmented and indistinguishable, whereas the expression levels of both proteins in UT5600 were similar to those in the C600 strain. This observation suggests that the difference in expression levels between the wild-type and  $\Delta$ SS proteins is due to degradation by the cytoplasmic Lon protease. This is not surprising, since the AIDA-I protein without the signal sequence should accumulate in the cytoplasm and should therefore be more susceptible to degradation by cytoplasmic proteases. Cellular fractionation was performed to confirm the location of the  $\Delta$ SS mutant and, as shown in Fig. 3B, the  $\Delta$ SS protein is indeed found in the cytoplasm, along with a lesser amount recovered in the membrane fraction.

### **Cytoplasmic glycosylation of AIDA-I**



We used a digoxigenin-hydrazine labeling kit to detect glycan and assess the glycosylation state of the  $\Delta$ SS protein. As shown in Fig. 4A, the  $\Delta$ SS protein is glycosylated. This result is in agreement with the observed molecular weight of approximately 100 kDa for the  $\Delta$ SS mature protein (Fig. 3A), which correspond to the glycosylated form of the mature protein whereas the unglycosylated form has a molecular weight of approximately 80 kDa [5, 8]. It is also consistent with the fact that the  $\Delta$ SS protein is recognized by antibodies directed against glycosylated mature AIDA-I (Fig. 3A) because we, and others, observed that these antibodies can only recognize the glycosylated form of the protein [5, 8].

These observations show that AIDA-I can be glycosylated in the cytoplasm. To confirm this result with a wild-type protein, bacteria expressing wild-type AIDA-I were grown in the presence of sodium azide in order to block Sec-dependent protein translocation. In those conditions, we observed an accumulation of the glycosylated pre-protein and proprotein, as detected with the antibodies raised against glycosylated mature AIDA-I (Fig. 4B). Taken together, these results suggest that the polypeptide can be synthesized and glycosylated in the cytoplasm without requirement for export across the inner membrane.

Since the  $\Delta$ SS protein is glycosylated, it can interact with Aah. It is therefore tempting to speculate that the amount of  $\Delta$ SS protein recovered in the membrane fraction, as seen in Fig. 3B, is due to a fraction of the protein remaining in association with the Aah glycosyltransferase at the membrane.

We also noted the presence of both proprotein and mature AIDA-I in whole cell extracts for the  $\Delta$ SS construct (Fig. 3A and Fig. 4C). This shows that the protein blocked in the cytoplasm can undergo proteolytic cleavage, despite the fact that this step is normally occurring after outer membrane translocation [28]. AIDA-I is cleaved by a unique mechanism among autotransporter, since the cleavage site is found in the extracellular domain of the protein and the protein is thought to cleave itself autocatalytically but does not bear any recognizable catalytic site [7, 28]. The fact that cleavage can occur in the cytoplasm, an abnormal location, is in agreement with the hypothesis that cleavage is autocatalytic [7, 28]. Proteolytic processing of the  $\Delta$ SS protein requires glycosylation, since the unglycosylated  $\Delta$ SS protein is not processed (Fig. 4C). We have previously shown that glycosylation is required for the normal conformation of the protein [8]. The cytoplasmic unglycosylated  $\Delta$ SS protein might be in a conformation incompatible with the processing, which is again in agreement with the idea that cleavage is autocatalytic.

**The N-terminal signal sequence extension is not essential for glycosylation.**

As explained above, all known autotransporters that are glycosylated possess a N-terminal signal sequence extension. This observation raises the possibility that this unusual signal sequence may be important for glycosylation. To test this hypothesis, we generated another signal sequence mutant of AIDA-I, SS $\Delta$ N, corresponding to a deletion of amino acids Lys3 to Leu22 (Fig. 1). The resulting protein possesses a typical signal sequence of 29 amino acids with the usual N, H and C domains.

Whole cell extracts of bacteria expressing wild-type AIDA-I, SS $\Delta$ N, or bearing a control empty vector were probed with antibodies against HisG tag or mature AIDA-I. As shown in Fig. 5A, the expression level of SS $\Delta$ N protein is considerably lower than for the wild-type protein. Previous studies have shown that the mature AIDA-I remains associated with the outer membrane by non-covalent interaction and can be released by treatment at 60°C [4]. As shown in Fig. 5B, the reduced amounts of SS $\Delta$ N protein can be heat-extracted and is therefore correctly located at the cell surface. A subcellular fractionation was also performed and confirmed that the low level of SS $\Delta$ N protein observed is located entirely in a membrane fraction (data not shown). We used a digoxigenin-hydrazine labeling kit to detect glycan and, as shown in Fig. 5C, the heat-extracted SS $\Delta$ N protein is glycosylated. This result shows that the N-terminal extension of the signal sequence is not essential for glycosylation.

The SS $\Delta$ N protein is produced in significantly less amount than the wild-type protein. The amount of transcripts coding for SS $\Delta$ N, measured by qRT-PCR, is 5-fold lower than the amount of transcript coding for wild-type AIDA-I and can only partially explain the important difference in protein levels between SS $\Delta$ N and wild-type AIDA-I. This difference could be due to additional effects of the deletion of the extension on translation. For instance, the truncation of the N-terminal extension of the unusually long signal sequence of *Bordetella pertussis* filamentous hemagglutinin results in a dramatic reduction in the rate of protein synthesis, as measured by radioactive labeling [21]. Alternatively, the extension might be involved in a later step of the biogenesis, such as the targeting to the Sec translocon and/or the kinetics of periplasmic release and folding, as was previously suggested [12, 18, 25, 29].

In conclusion, our results show that the SAATs-specific heptosyltransferase Aah is a membrane-associated protein able to glycosylate AIDA-I in the bacterial cytoplasm,

independently of the export of the latter through the Sec translocon. Additionally, we have shown that the unusual N-terminal extension of the Sec-dependent signal sequence of AIDA-I is not essential for glycosylation. Cytoplasmic glycosylation of AIDA-I was postulated before [5], but not proven. Cytoplasmic glycosylation of bacterial protein is rare in Gram-negative bacteria. Interestingly, the other instance concerns the HMW1 adhesin of *Haemophilus influenzae*, which is also secreted by the type V pathway [14]. HMW1 is specifically modified by HMW1C and the resulting glycan structure contains galactose, glucose and mannose. As with Aah and TibC, however, the mechanism by which HMW1C mediate glycosylation remains unknown.

The association of Aah with membranes was not hypothesized previously. It is not unexpected, however, because Aah shares homologies with the *E. coli* LPS heptosyltransferases, proteins which are known to be associated with the inner membrane and face the cytoplasm [19].

Our results suggest that proteins glycosylated in the cytoplasm can then be secreted. This raises several important issues. The modification could for instance interfere with Sec-dependent translocation and therefore require specific accessory proteins. The requirement for such accessory proteins has indeed been observed for the secretion of the cytoplasmically glycosylated GspB adhesin of *S. gordonii* [2]. Cytoplasmic glycosylation of a secreted polypeptide also implies a coordination between translation, glycosylation and secretion. It is worth noting that we do not know if glycosylation occurs co- or post-translationally. Lastly, the cytoplasmic glycosylation suggests that the substrate of Aah is an unfolded polypeptide. The specificity of Aah is then mysterious since a consensus sequence for the modification it performs has not been identified [8, 27]. These questions warrant further studies of the unique unusual mechanism of action of the Aah protein glycosyltransferase.

### **Acknowledgements**

This work was supported by financial contributions from the Groupe de Recherche et d'Etudes sur les Maladies Infectieuses du Porc (GREMIP), the Natural Sciences and Engineering Research Council of Canada (NSERC # 262746), the Canadian Institutes of Health Research (CIHR # 84578), the Canada Research Chair program and the Canada Foundation for Innovation (# 201414), and a graduate fellowship to M.-E. C. from the Natural Sciences and Engineering Research Council of Canada (NSERC # 85297596).

## References

- [1] Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 340, 783-795.
- [2] Bensing, B.A., Takamatsu, D., Sullam, P.M. (2005) Determinants of the *streptococcal* surface glycoprotein GspB that facilitate export by the accessory Sec system. *Mol Microbiol.* 58, 1468-1481.
- [3] Benz, I., Schmidt, M.A. (1989) Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infect Immun.* 57, 1506-1511.
- [4] Benz, I., Schmidt, M.A. (1992) AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. *Mol Microbiol.* 6, 1539-1546.
- [5] Benz, I., Schmidt, M.A. (2001) Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin. *Mol Microbiol.* 40, 1403-1413.
- [6] Chapman, T.A., Wu, X.Y., Barchia, I., Bettelheim, K.A., Driesen, S., Trott, D., Wilson, M., Chin, J.J. (2006) Comparison of virulence gene profiles of *Escherichia coli* strains isolated from healthy and diarrheic swine. *Appl Environ Microbiol.* 72, 4782-4795.
- [7] Charbonneau, M.E., Berthiaume, F., Mourez, M. (2006) Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I). *J Bacteriol.* 188, 8504-8512.
- [8] Charbonneau, M.E., Girard, V., Nikolakakis, A., Campos, M., Berthiaume, F., Dumas, F., Lepine, F., Mourez, M. (2007) O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence. *J Bacteriol.* 189, 8880-8889.
- [9] Chavan, M., Lennarz, W. (2006) The molecular basis of coupling of translocation and N-glycosylation. *Trends Biochem Sci.* 31, 17-20.
- [10] Chevalier, N., Moser, M., Koch, H.G., Schimz, K.L., Willery, E., Loch, C., Jacob-Dubuisson, F., Muller, M. (2004) Membrane targeting of a bacterial virulence factor harbouring an extended signal peptide. *J Mol Microbiol Biotechnol.* 8, 7-18.
- [11] Desvaux, M., Cooper, L.M., Filenko, N.A., Scott-Tucker, A., Turner, S.M., Cole, J.A., Henderson, I.R. (2006) The unusual extended signal peptide region of the type V secretion system is phylogenetically restricted. *FEMS Microbiol Lett.* 264, 22-30.

- [12] Desvaux, M., Scott-Tucker, A., Turner, S.M., Cooper, L.M., Huber, D., Nataro, J.P., Henderson, I.R. (2007) A conserved extended signal peptide region directs posttranslational protein translocation via a novel mechanism. *Microbiology*. 153, 59-70.
- [13] Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M., Brinkman, F.S. (2005) PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*. 21, 617-623.
- [14] Grass, S., Buscher, A.Z., Swords, W.E., Apicella, M.A., Barenkamp, S.J., Ozchlewski, N., St Geme, J.W. (2003) The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol.Microbiol.* 48, 737-751.
- [15] Henderson, I.R., Nataro, J.P. (2001) Virulence functions of autotransporter proteins. *Infect Immun.* 69, 1231-1243.
- [16] Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C., Ala'Aldeen, D. (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev.* 68, 692-744.
- [17] Henderson, I.R., Navarro-Garcia, F., Nataro, J.P. (1998) The great escape: structure and function of the autotransporter proteins. *Trends Microbiol.* 6, 370-378.
- [18] Jong, W.S., Luirink, J. (2008) The conserved extension of the Hbp autotransporter signal peptide does not determine targeting pathway specificity. *Biochem Biophys Res Commun.*
- [19] Kadmas, J.L., Raetz, C.R. (1998) Enzymatic synthesis of lipopolysaccharide in *Escherichia coli*. Purification and properties of heptosyltransferase I. *J Biol Chem.* 273, 2799-2807.
- [20] Klemm, P., Vejborg, R.M., Sherlock, O. (2006) Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol.* 296, 187-195.
- [21] Lambert-Buisine, C., Willery, E., Locht, C., Jacob-Dubuisson, F. (1998) N-terminal characterization of the *Bordetella pertussis* filamentous haemagglutinin. *Mol Microbiol.* 28, 1283-1293.
- [22] Lindenthal, C., Elsinghorst, E.A. (1999) Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect Immun.* 67, 4084-4091.
- [23] Moormann, C., Benz, I., Schmidt, M.A. (2002) Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infect Immun.* 70, 2264-2270.

- [24] Mourez, M., Hofnung, M., Dassa, E. (1997) Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J.* 16, 3066-3077.
- [25] Peterson, J.H., Szabady, R.L., Bernstein, H.D. (2006) An unusual signal peptide extension inhibits the binding of bacterial presecretory proteins to the signal recognition particle, trigger factor, and the SecYEG complex. *J Biol Chem.* 281, 9038-9048.
- [26] Rutherford, N., Charbonneau, M.E., Berthiaume, F., Betton, J.M., Mourez, M. (2006) The periplasmic folding of a cysteineless autotransporter passenger domain interferes with its outer membrane translocation. *J Bacteriol.* 188, 4111-4116.
- [27] Sherlock, O., Dobrindt, U., Jensen, J.B., Munk Vejborg, R., Klemm, P. (2006) Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. *J Bacteriol.* 188, 1798-1807.
- [28] Suhr, M., Benz, I., Schmidt, M.A. (1996) Processing of the AIDA-I precursor: removal of AIDA<sub>c</sub> and evidence for the outer membrane anchoring as a beta-barrel structure. *Mol Microbiol.* 22, 31-42.
- [29] Szabady, R.L., Peterson, J.H., Skillman, K.M., Bernstein, H.D. (2005) An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proc Natl Acad Sci U S A.* 102, 221-226.
- [30] Szymanski, C.M., Wren, B.W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol.* 3, 225-237.
- [31] VanderVen, B.C., Harder, J.D., Crick, D.C., Belisle, J.T. (2005) Export-mediated assembly of *mycobacterial* glycoproteins parallels eukaryotic pathways. *Science.* 309, 941-943.
- [32] Wu, H., Bu, S., Newell, P., Chen, Q., Fives-Taylor, P. (2007) Two gene determinants are differentially involved in the biogenesis of FapI precursors in *Streptococcus parasanguis*. *J Bacteriol.* 189, 1390-1398.
- [33] Zhang, W., Zhao, M., Ruesch, L., Omot, A., Francis, D. (2007) Prevalence of virulence genes in *Escherichia coli* strains recently isolated from young pigs with diarrhea in the US. *Vet Microbiol.* 123, 145-152.

### Figures legends

**Fig. 1: Schematic representation of the signal sequence mutants of AIDA-I.** The three domains of the AIDA-I autotransporter are shown: the N-terminal signal sequence (SS), the

central domain (Mature AIDA-I) and the C-terminal domain (AIDAc). The black box represents the location of a HisG tag at the N-terminus of the proprotein. The sequence of the 49 amino acids signal peptide is shown, with its N-terminal extension, basic N domain, hydrophobic H domain and C-terminal domain. The sequences of the deletion mutants are given.

**Fig. 2: Subcellular location of the Aah glycosyltransferase.** *A.* Periplasmic (PER), cytoplasmic (CYT) and insoluble (INS) fractions of a subcellular fractionation of cultures of *E. coli* strain BL21 bearing the empty vector or the pAah plasmid were separated by SDS-PAGE and stained with Coomassie blue. The black arrowhead represents the Aah protein, as confirmed by N-terminal sequencing. *B.* A culture of BL21 bearing pAahH, a plasmid allowing the expression of His-tagged Aah, was fractionated similarly but the insoluble material was then treated sequentially with 1% Triton X-100 and 8M Urea. Thus, the following fractions were collected and probed with an antibody against the His-tag: periplasm (PER), cytoplasm (CYT), Triton X-100 soluble (T<sub>S</sub>), Triton X-100 insoluble-urea soluble (U<sub>S</sub>) and Triton X-100/urea insoluble (T<sub>I</sub>/U<sub>I</sub>).

**Fig. 3: Expression and location of the ΔSS mutant of AIDA-I.** *A.* Whole cell extracts were obtained from overnight cultures of *E. coli* C600, BL21 or UT5600 harboring an empty vector (-), plasmids pAngH and pAah (WT) or pAngHΔSS and pAah (ΔSS). Proteins separated by SDS-PAGE were probed by immunoblotting with antibodies directed against mature AIDA-I or with an antibody directed against the HisG tag. This allows detection of the proprotein (Pro., 145 kDa) and mature AIDA-I (100 kDa). *B.* Subcellular fractionation of cultures of C600 bearing the same plasmids. The following fractions were collected and probed with an antibody against the HisG tag: cytoplasm (CYT), total membranes (M) and periplasm (PER).

**Fig. 4: Cytoplasmic glycosylation of AIDA-I.** *A.* Detection of glycosylation, using digoxigenin-hydrazine glycan labeling, in whole cell extracts of *E. coli* BL21 cultures harboring empty vector (-), plasmids pAngH and pAah (WT) or pAngHΔSS and pAah (ΔSS). *B.* Cultures of *E. coli* C600 harboring plasmids pAngH and pAah (WT) were grown in presence or absence of sodium azide 0.02% (NaN<sub>3</sub>). Proteins from whole cell extracts were separated and probed with antibodies directed against mature AIDA-I, which only

recognizes the glycosylated protein [8]. In these conditions the pre-protein (Pre.) is visible. C. Proteins from whole cell extracts of overnight cultures of *E. coli* BL21 harboring plasmids pAngH (WT) or pAngH $\Delta$ SS ( $\Delta$ SS), with or without plasmid pAah, were probed with an antibody against the HisG tag.

**Fig. 5: The signal sequence N-terminal extension is not essential for glycosylation.** A. Whole cell extracts were obtained from overnight cultures of *E. coli* C600 harboring an empty vector (-), plasmids pAngH and pAah (WT) or pAngHSS $\Delta$ N and pAah (SS $\Delta$ N). Proteins separated by SDS-PAGE were probed by immunoblotting with an antiserum against the HisG tag or with an antibody against mature AIDA-I. B. Heat extracts of these cultures were separated by SDS-PAGE and stained with Coomassie blue. C. Detection of glycosylation, using digoxigenin-hydrazine glycan labeling, in the heat-extracts.



Figure 1

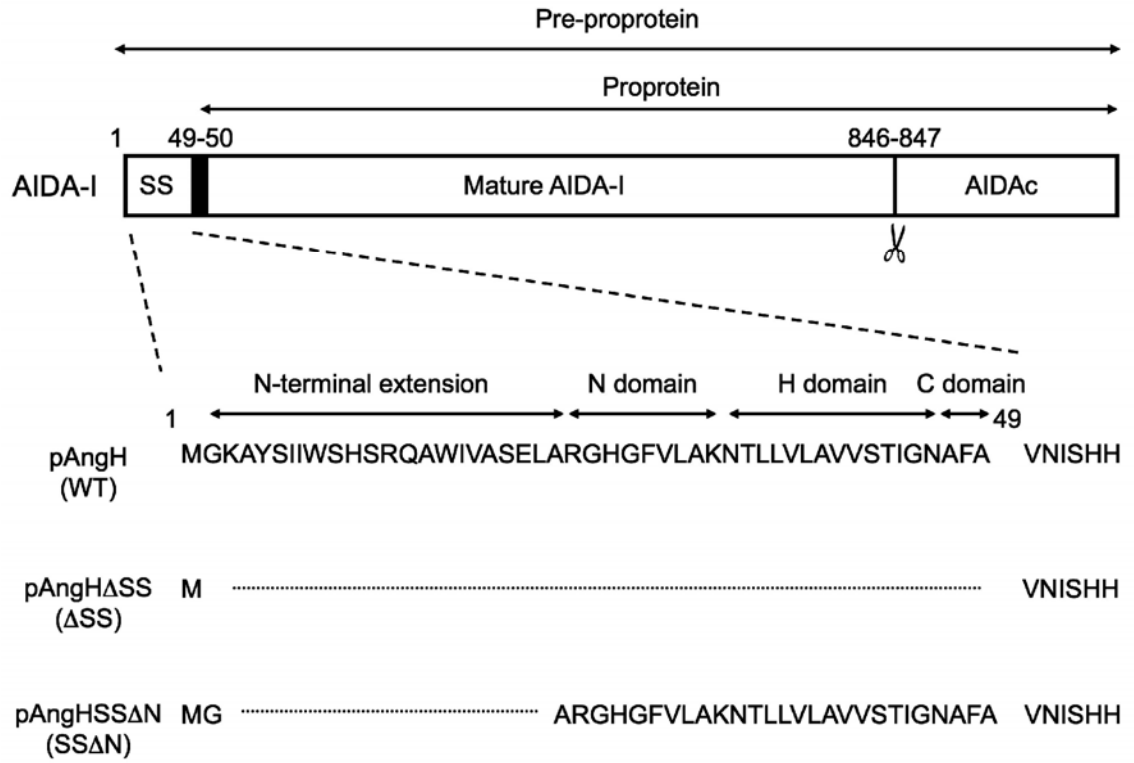


Figure 2

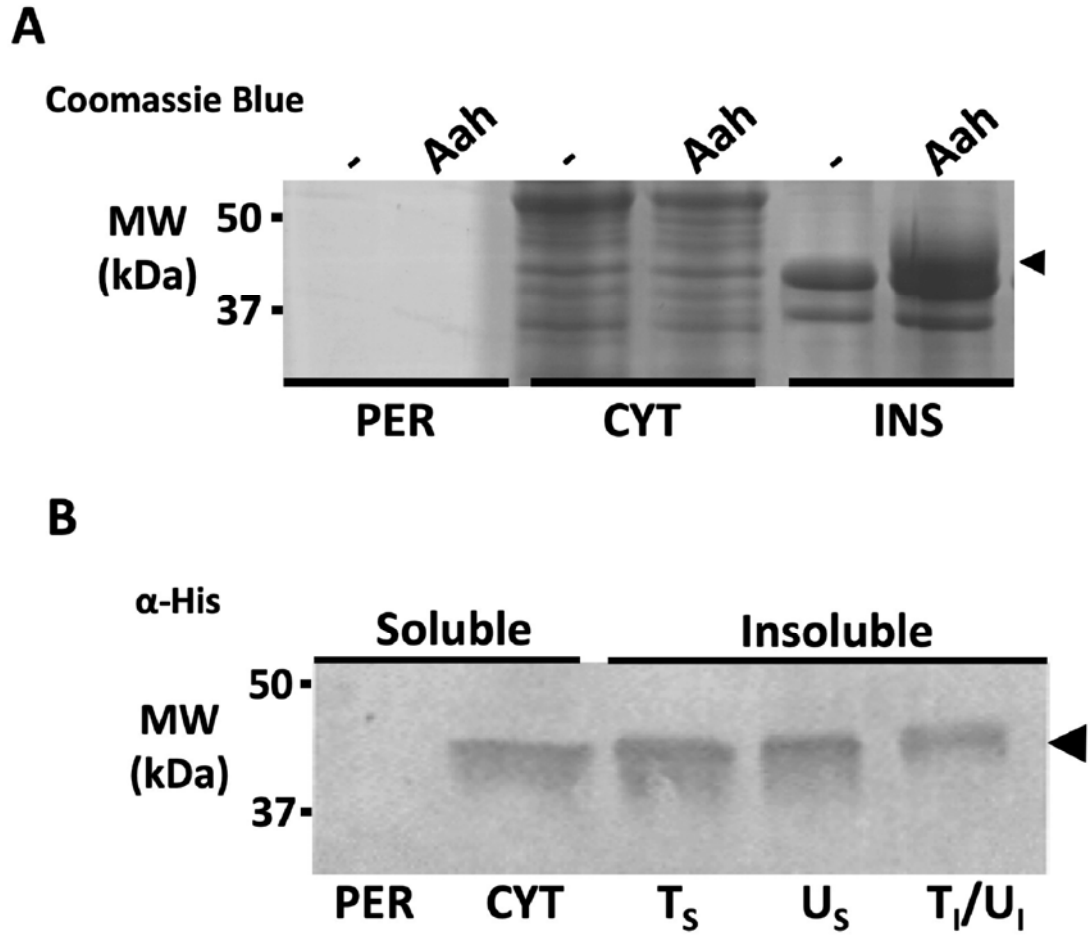


Figure 3

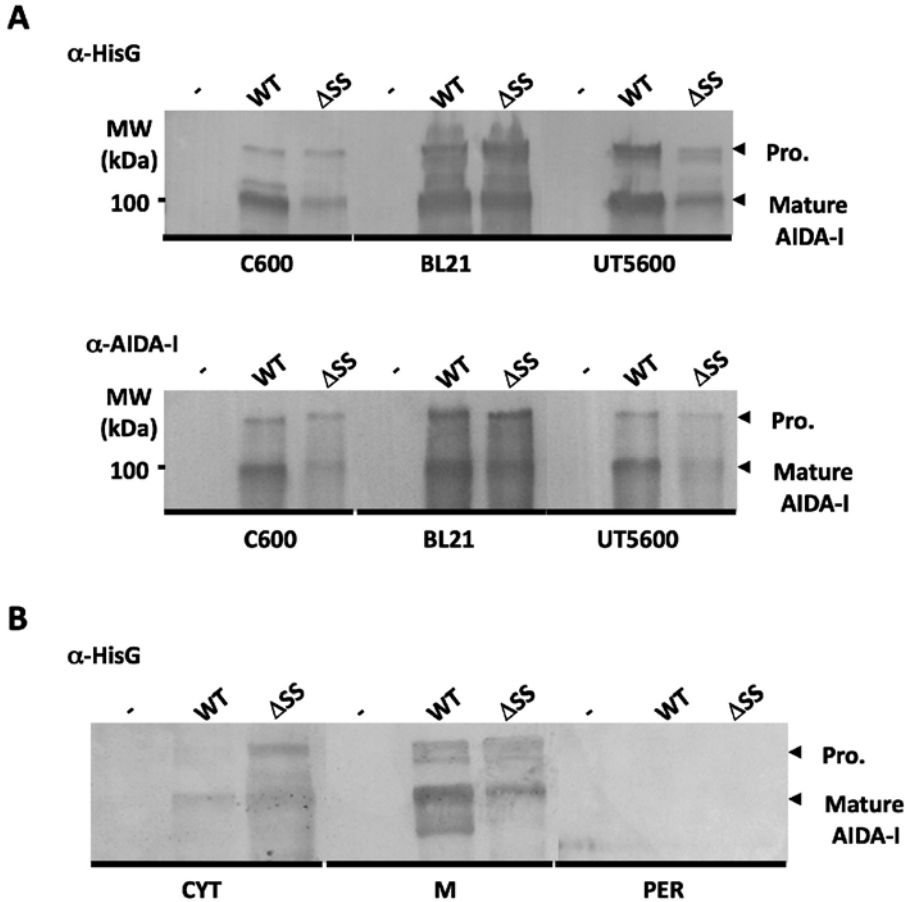


Figure 4

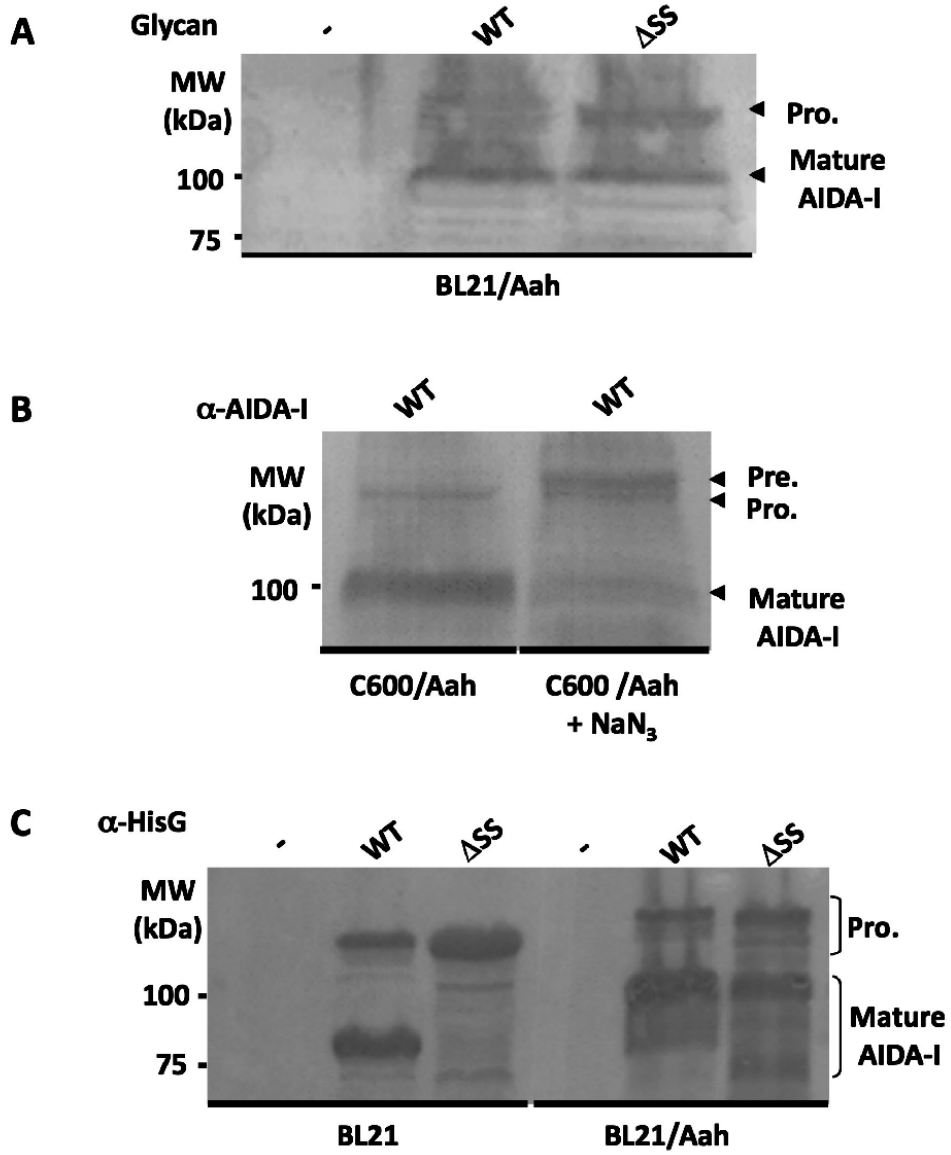
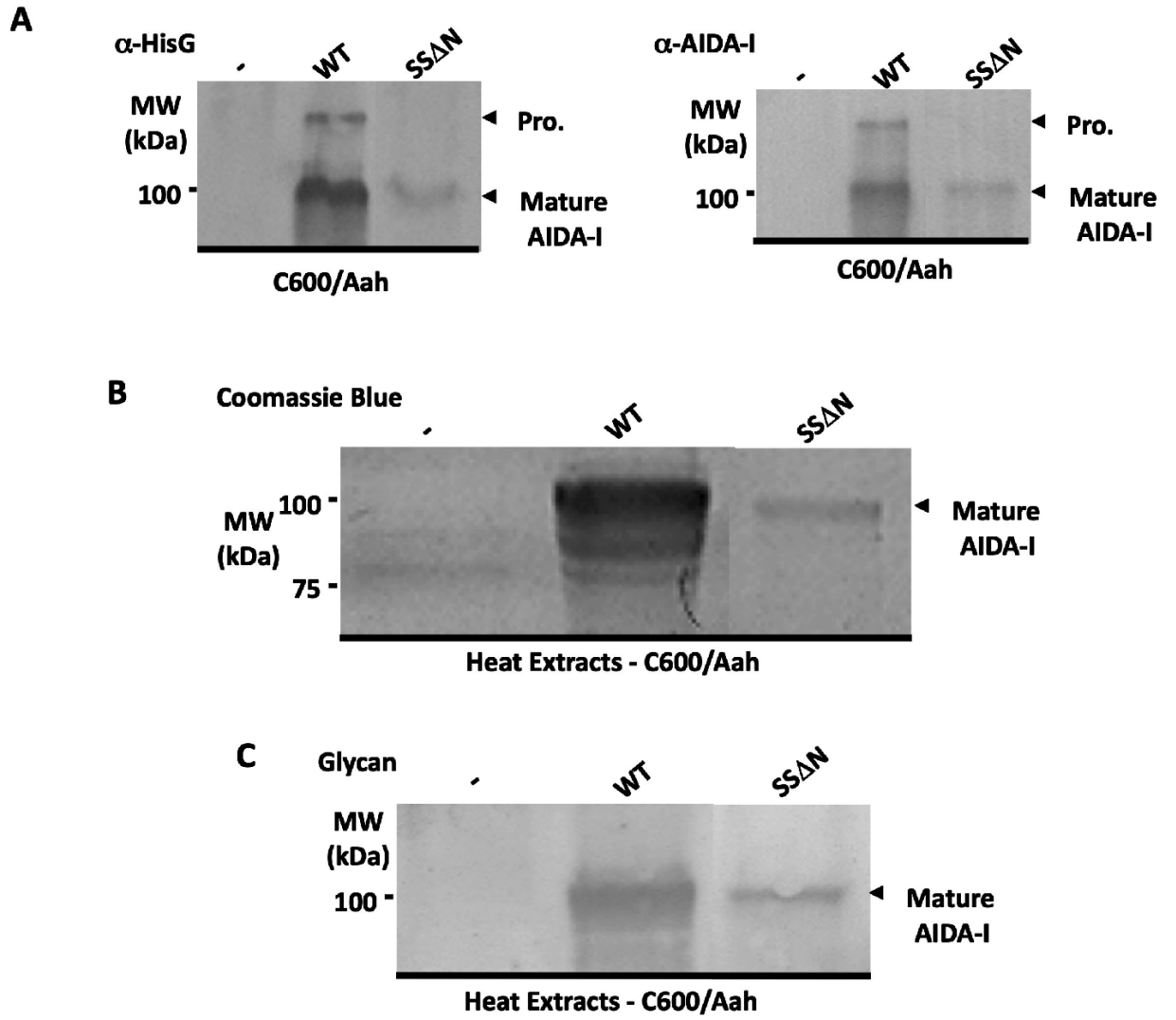


Figure 5



## ARTICLE 5

**Autoprocessing of the *Escherichia coli* AIDA-I autotransporter : a new mechanism involving acidic residues in the junction region**

Journal of Biological Chemistry, 2009, Jun., Volume 284 (25), pages 17340-51

## Contribution de l'auteur pour cet article

En tant que premier auteur de cet article, j'ai effectué la plupart des expérimentations décrites, à l'exception de quelques mutants ponctuels. J'ai aussi écrit la totalité de l'article.

## Contribution scientifique de l'article

Dans cet article, nous avons décrit un nouveau mécanisme de clivage autocatalytique de nature intramoléculaire pour la famille des autotransporteurs monomériques.

**Autoprocessing of the *Escherichia coli* AIDA-I autotransporter; a new mechanism involving acidic residues in the junction region**

Marie-Ève Charbonneau, Julie Janvire and Michael Mourez

Running title: Autoprocessing of the AIDA-I autotransporter

From the Canada research chair on bacterial animal diseases, Université de Montréal, Faculté de Médecine Vétérinaire, 3200 sicotte, St-Hyacinthe, J2S 7C6, Québec, Canada

**Abstract**

The cleavage of the autotransporter adhesin involved in diffuse adherence (AIDA-I) of *Escherichia coli* yields a membrane-embedded fragment, AIDAc, and an extracellular fragment, the mature AIDA-I adhesin. The latter remains noncovalently associated with AIDAc, but can be released by heat treatment. In this study, we determined the mechanism of AIDA-I cleavage. We showed that AIDA-I processing is an auto-catalytic event by monitoring the *in vitro* cleavage of an uncleaved mutant protein isolated from inclusion bodies. Furthermore, by following changes in far UV CD spectra and protease resistance of the renatured protein, we showed that the cleavage of the protein is correlated with folding. By site-directed deletions, we showed that the catalytic activity of the protein lies in a region encompassing amino acids between A667 and T953, which includes the conserved junction domain of some autotransporter. With site-directed point mutations, we also found that D878 and E897 are involved in the processing of AIDA-I and that mutation preserving the acidic side chain of D878 was tolerated, giving evidence that this carboxylic acid group is directly involved in catalysis. Lastly, we confirmed that cleavage of AIDA-I is intramolecular. Our results unveil a new mechanism of auto-processing in the autotransporter family.



## Introduction

Monomeric autotransporters, secreted by the type Va secretion pathway, constitute one of the largest family of secreted proteins in Gram-negative bacteria (1). Various virulence attributes have been associated with most autotransporters, such as adhesion or invasion, self-association, biofilm formation, serum resistance, cytotoxic activity, to name only a few (2,3). Autotransporters are synthesized as pre-proteins with modular organization. An N-terminal *sec*-dependent signal sequence permits the secretion of the protein across the inner membrane (4). A C-terminal membrane-embedded domain promotes secretion of parts of the protein across the outer membrane and is composed of a  $\beta$ -barrel and a  $\alpha$ -helix spanning its lumen (5,6). The central domain of the protein is thus extracellular and bears its functional part. In some cases, the extracellular part of the protein is cleaved and remains associated with the outer membrane or is secreted in the extracellular milieu. Just before the membrane-embedded domain, a functional subdomain of approximately 100 amino acids is sometimes present in the extracellular domain and has been called the junction region (also named autochaperone domain or stable core) (7-9). This region is essential for stabilizing the  $\beta$ -barrel and/or to promote folding of the extracellular domain (7-9).

The adhesin involved in diffuse adherence (AIDA-I) is one monomeric autotransporter that has been extensively studied. AIDA-I was originally identified as a plasmid-encoded protein from a diffusely adhering *Escherichia coli* strain isolated in a case of infantile diarrhea (10). This adhesin was then shown to play a role in neonatal and postweaning diarrheal diseases in piglets, both of which cause major economic losses in farms worldwide (11-13). Besides its role as an adhesin, AIDA-I has been shown to mediate self-association and biofilm formation (14) as well as invasion of epithelial cells (15). Additionally, this protein undergoes a modification that is rare in bacteria, as it is O-glycosylated by the specific cytoplasmic protein autotransporter adhesin heptosyltransferase (Aah) (16,17). AIDA-I has been proposed to be a member of a new group of autotransporter called *self-associating autotransporters* (SAATs), which includes the Ag43 aggregation factor and the TibA invasin (18).

After secretion at the cell surface, the extracellular domain of the protein, called mature AIDA-I, is cleaved from the C-terminal domain, called AIDA<sub>c</sub>, which encompasses the  $\beta$ -barrel, the  $\alpha$ -helix and the junction region (19,20). Mature AIDA-I, however, remains strongly associated with AIDA<sub>c</sub> (15). The processing of AIDA-I is essential neither for the

multiple functions of the protein, nor for its stability (15). Questions remain about the mechanism involved in this cleavage. In a few autotransporters, the cleavage reaction involves an exogenous protease. IcsA for instance, an autotransporter of *Shigella Flexneri*, is cleaved by a dedicated outer membrane protease called IcsP, which is related to the OmpT protease found in *E. coli* (21). Similarly, the *Neisseria meningitidis* serine protease NalP is specifically responsible for the partial processing of various autotransporters including the IgA protease (22), App (22), AusI (23) and MspA (24). In contrast, the cleavage of AIDA-I is not mediated by the known *E. coli* periplasmic or outer membrane proteases (DegP, OmpT or OmpP), and it occurs in *E. coli* as well as in *Shigella* or *Salmonella* strains (19). Based on those observations, this processing has been suggested to be an autocatalytic event, even though no protease catalytic site can be identified in the protein.

Two types of autocatalytic processing mechanisms have been described in other autotransporters. The Hap autotransporter of *Haemophilus influenzae* (25), the SphB1 protein of *Bordetella pertussis* (26) and the NalP (22), App (27) and IgA proteases of *N. meningitidis* (28) are cleaving themselves by an endogenous serine protease domain. More recently, it was shown that the serine protease autotransporters of Enterobacteriaceae (SPATE) family are cleaving themselves in the  $\alpha$ -helix located in the lumen of the membrane-embedded domain by a catalytic reaction that involves an aspartate and a conserved asparagine (6,29). The same autocatalytic reaction has been shown for two distantly related *B. pertussis* autotransporters, pertactin and BrkA (29). However, none of these mechanisms can be hypothesized for AIDA-I cleavage: AIDA-I possesses neither a serine protease domain like Hap, nor the asparagine residue required for the SPATE cleavage. In addition, the AIDA-I cleavage site is not located in the  $\alpha$ -helix. Thus, the mechanism of AIDA-I processing must be different and remains unknown.

The mechanisms by which a diversity of other autotransporters are cleaved also remain unknown. This is for instance the case of the *E. coli* Ag43 SAAT (30), the *Helicobacter pylori* VacA cytotoxin (31), the *Chlamydia pneumonia* polymorphic membrane proteins (32), the Ssp proteins of *Serratia marcescens* (33) as well as the cohemolysin Cfa (34) and the Arp immunogenic protein of *Bartonella Henselae* (35).

In this study, we determined the mechanism of AIDA-I cleavage. We first showed that a truncated mutant of AIDA-I can refold *in vitro* and undergo self-cleavage, providing definite proof for the autocatalytic mechanism. By constructing site-directed deletions and

point mutations, we then showed that the catalytic domain lies in the region encompassing the amino acids between A667 and T953, and we identified two residues in the junction region, D878 and E897, that are essential for processing. Lastly, we confirmed that the mechanism of proteolysis for AIDA-I is intramolecular. This is the first report of this type of autocatalytic activity.

### **Experimental procedures**

**Bacterial strains and plasmids.** The *Escherichia coli* K12 strains C600 (obtained from New England Biolabs, genotype: F<sup>-</sup>, *thr1*, *leuB6*, *thi1*, *lacY1*, *supE44*, *rfdD1*, *fhuA21*) and BL21 (obtained from Stratagene, genotype: F<sup>-</sup>, *ompT*, *hsdS*(r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>), *gal*) were used in this study. Plasmid pAgH has been described before and allows expression of Aah, and of AIDA-I tagged at the N-terminus of the proprotein with six histidine amino acids and a glycine (HisG) (15). All point mutations in the pAgH plasmids were introduced by oligonucleotide-directed mutagenesis with the QuickChange II site-directed mutagenesis kit (Stratagene) using the primers listed in Table 1 and their corresponding complementary oligonucleotides. The pAngH (allowing expression of AIDA-I without Aah) pAngHΔSS (deletion of amino acids 1 to 49), pAngHΔN (deletion of amino acids 54 to 225) and pAngHΔC2 (deletion of amino acids 224 to 667) plasmids were described elsewhere (36,37). The Δαβ deletion (deletion of amino acids 953 to 1286) was introduced into pAngHΔSS also by site-directed mutagenesis using primers listed in Table 1, generating the pAngHΔSSΔαβ plasmid. For glycosylation, we used plasmid pAah, a derivative of pACYC-184 described elsewhere (17). Plasmid pACYC-AgH was obtained by sub-cloning in the pAah vector the DNA fragments from pAgH generated by restriction enzymes XbaI and BrsGI.

**Functional assays.** The functionality of the uncleaved mutant pAgHD878N was determined by the functional assays for adhesion, autoaggregation and biofilm formation, as previously described (15).

**Bacteria growth and induction conditions.** Bacteria containing the different plasmids were grown on Luria-Bertani (LB) agar plates or in LB broth containing 100 μg.ml<sup>-1</sup> ampicillin and, when required, 50 μg.ml<sup>-1</sup> chloramphenicol. Bacterial cultures were grown at 30°C and induced with 10 μM isopropyl-β-thiogalactopyranoside (IPTG) when their

reached an optical density at 600 nm ( $OD_{600nm}$ ) of 0.8, unless indicated otherwise. This low concentration of IPTG was used to limit the toxicity associated with overexpression of AIDA-I. To produce inclusion bodies, bacteria were grown at 37°C until an  $OD_{600nm}$  of 0.8 and induced overnight with 1mM IPTG.

**SDS-PAGE and immunoblotting.** Protein samples were diluted in twice-concentrated SDS-PAGE loading buffer containing  $\beta$ -mercaptoethanol, and denatured by heating at 100°C for 10 minutes. The samples were then separated by SDS-PAGE on 10% acrylamide gels. The gels were either stained with Coomassie blue or transferred to polyvinylidene fluoride membranes (Millipore). Immunodetection was performed with anti-His antibodies (Applied biological materials) diluted 1:10,000 in blocking buffer (5% skim milk, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Triton X-100). A rabbit anti-mouse horseradish peroxidase-conjugated antibody (Sigma) was used as a secondary antibody according to the instructions of the manufacturer. Immune complexes were revealed using a 3,3',5,5'-Tetramethylbenzidine solution for membranes (Sigma).

**Whole cell extracts.** Overnight cultures were grown, normalized and centrifuged for 15 minutes at 12,000 x g in microfuge tubes and the pellets were resuspended in 100  $\mu$ l of Phosphate-Buffered Saline (PBS). All samples were processed by SDS-PAGE and immunoblotting, as described above.

**Analysis of the intra- or intermolecular nature of AIDA-I processing.** Overnight cultures of *E. coli* C600 harboring plasmid pTRC99a, pAgH or pAgHD878N were normalized at the same  $OD_{600nm}$  and co-incubated for 30 minutes at 30°C. The samples were pelleted and whole-cell extracts were obtained as described above. Alternatively, whole cell extracts of overnight cultures *E. coli* C600 harboring plasmids pTRC99a/PACYC184, pACYC-AgH, pAgHD878N or pACYC-AgH and pAgHD878N simultaneously, were prepared as described above. For the *in vitro* experiment, native AIDA-I proteins (wild-type and D878N mutant) were purified as described below. Pure proteins were incubated separately or together for 30 minutes at room temperature or for 2 hours at 4°C, resolved by SDS-PAGE and the gels stained with Coomassie blue.

**Purification of native AIDA-I.** The purification was performed as described previously

(15). Briefly, one liter of overnight cultures of C600 harboring plasmids pAgH or pAgHD878N were grown until an  $OD_{600nm}$  of 0.8 and induced with 10  $\mu$ M of IPTG. Bacteria were harvested and resuspended in 50 ml of Tris-Buffered Saline (TBS: 50 mM Tris-HCl pH 8, 150 mM NaCl) containing lysozyme (0.4 mg.ml<sup>-1</sup> final concentration) and EDTA pH8 (10 mM final concentration) and lysed with French press and an ultrasonic processor. Cellular debris were removed by low speed centrifugation. The lysate was centrifuged for 1 hour in an ultracentrifuge at 250,000  $\times$  g. The membranes were resuspended in TBS containing 1% Triton X-100, incubated for 1 hour at room temperature, and centrifuged again. The solubilized membranes contained the histidine-tagged native AIDA-I. The proteins were purified using an ÄKTA purifier system with a 1-ml His Trap HP column (Amersham biosciences), according to the instructions of the manufacturer. The purity of the purified proteins was confirmed by SDS-PAGE and staining with Coomassie blue.

**Purification of inclusion bodies and refolding.** – One liter of *E. coli* BL21 harboring plasmid pAngH $\Delta$ SS $\Delta$  $\alpha\beta$  or pAngH $\Delta$ SS $\Delta$  $\alpha\beta$ D878N and plasmid pAah were grown at 37°C until an  $OD_{600nm}$  of 0.8 and induced overnight with 1 mM IPTG to allow formation of inclusion bodies. Bacteria were harvested and resuspended in 50 ml of TBS containing a proteases inhibitor cocktail (Complete EDTA-free, Roche) and lysed with French press and an ultrasonic processor. Cellular debris were removed by low speed centrifugation. Inclusion bodies were collected by a centrifugation at 4,000  $\times$  g for 75 minutes. The pellets were resuspended in 30 ml of denaturing buffer (0.1 M Tris-HCl pH8, 6 M guanidinium hydrochloride) and centrifuged again at 16,000  $\times$  g for 1 hour to remove any debris. Denaturated proteins were purified using an ÄKTA purifier system with a 1-ml His Trap HP column, as above.

Renaturation of the proteins was performed by a progressive replacement of the denaturing buffer with the renaturation buffer (TBS pH8, 0.1 % Triton X-100, 100  $\mu$ M phenylmethanesulphonyl fluoride (PMSF) and Complete EDTA-free proteases inhibitor cocktail) on a 1-ml His trap HP column for 2 hours. The proteins were then kept at 4°C for renaturation to proceed.

Alternatively, to study the effect of pH on the *in vitro* folding of AIDA-I, renaturation was performed by a progressive replacement of the denaturing agent by a buffer containing 2 mM Tris-HCl pH8, 150 mM NaCl, 0.1 % Triton X-100, 100  $\mu$ M PMSF and Complete

EDTA-free proteases inhibitor cocktail on a 1-ml His trap HP column for 2 hours. An equal amount of protein was diluted to a final concentration of 250  $\mu\text{g/ml}$  in buffers with different pH value supplemented with 150 mM NaCl, 0.1% Triton X-100, 100  $\mu\text{M}$  PMSF and Complete EDTA-free (50 mM sodium citrate, pH4 and pH6; 50 mM Tris-HCl, pH8; 50 mM sodium carbonate, pH10) and left at 4°C. For the pH shift experiment, Tris base 1M pH13 was used to adjust the pH4 buffer to pH8.

For mass spectrometry experiment, the proteins (native or refolded *in vitro*) were reduced prior to in gel digestion with trypsin. Tryptic peptides were separated with a SB-C18 ZORBAX 300 reverse-phase column and mass spectra were recorded on a hybrid linear ion trap-triple quadrupole mass spectrometer (Q-Trap; AB Applied Biosystems) equipped with a nanoelectrospray ionization source (F. Lépine, IAF-INRS, Laval, Canada).

**Limited proteolytic digestion.** Pure native proteins (wild-type or D878N mutant) were incubated on ice for 30 minutes in the presence of different concentrations of trypsin, varying between 167 and 1  $\mu\text{g.ml}^{-1}$  (final concentration). Alternatively, the renatured proteins were incubated on ice for 30 minutes in the presence or absence of trypsin at a final concentration of 37  $\mu\text{g.ml}^{-1}$ . Proteolysis was stopped by boiling samples for 10 minutes in SDS-PAGE loading buffer. All samples were processed by SDS-PAGE and Coomassie blue staining, as described above.

**Far-UV CD spectroscopy.** Far-UV circular dichroism (CD) spectra were recorded in a 0.1-cm-pathlength cuvette between 205 and 260 nm, using a spectropolarimeter (model J-810; Jasco spectroscopic Co. Ltd.). For each spectrum, fifteen accumulations were averaged. Thermal denaturation was followed by monitoring the ellipticity at 218 nm with the temperature varying between 25°C and 80°C at a rate of 5°C per min. Proteins concentration for all CD experiments was 190  $\mu\text{g.ml}^{-1}$  in TBS containing 0.1% Triton X-100. Ellipticities were converted to mean residual ellipticities.

**Structure prediction and alignment.** The three-dimensional structural model of the junction region and the cleavage site of AIDA-I (amino acids I807-L956) was constructed using the Phyre program (<http://www.sbg.bio.ic.ac.uk/phyre/>) (38) and is based on the known crystal structure of *B. pertussis* pertactin (39). Cartoon representation was generated using the PyMol software (Delano scientific LLC). Sequence alignment was performed

using the BLOSUM 62 matrix, with the Geneious Pro 4.0.2 software (Biomatters Ltd.).

## Results

**A purified mutant of AIDA-I undergoes self-cleavage *in vitro*.** As AIDA-I cleavage is thought to be autocatalytic, we decided to purify the unfolded glycosylated proprotein and monitor the *in vitro* folding and apparition of cleavage. To achieve this goal, we first used a previously described variant of AIDA-I without signal sequence that remains confined in the cytoplasm (36). This mutant was co-expressed with the glycosyltransferase Aah, as glycosylation is important to ensure the normal conformation of the protein (17). At high level of induction with IPTG, the AIDA-I proprotein forms inclusion bodies (data not shown). The inclusion bodies can be solubilized with 6 M guanidinium hydrochloride (GnHCl) and the uncleaved form of the protein can be purified by affinity chromatography with the help of an N-terminal 6 histidines tag. We were, however, unable to refold this protein, even when we tried a broad spectrum of buffers and conditions (data not shown). We hypothesized that the presence of the  $\beta$ -barrel is incompatible with correct folding of the entire protein *in vitro*. To solve this problem, we used a mutant of the signal sequence-less version of AIDA-I, which we called AIDA-I\* and lacks the  $\alpha$ -helix and the  $\beta$ -barrel but can still be cleaved (Fig.1A).

The uncleaved AIDA-I\* protein was solubilized from inclusion bodies in 6 M GnHCl and purified (data not shown). Renaturation was performed by exchanging GnHCl with a renaturation buffer on a 1-ml His trap HP column for 2 hours. Cysteine and serine protease inhibitors were added to the renaturation buffer to limit the risk of an unspecific cleavage being mediated by contaminating proteases. As show in Fig.1B, mature AIDA-I\* becomes detectable 43 hours after buffer exchange and increases in amount throughout the experiment. Furthermore, the appearance of the mature AIDA-I\* is concomitant with a reduction in the amount of proprotein. The time course of mature AIDA-I\* appearance could be fitted to single exponential association with a rate constant of  $0.53 \pm 0.19 \text{ min}^{-1}$  (Fig.1B). We used a 16 % tris-tricine gel to visualize the C-terminal peptide that would result from the cleavage of the proprotein into mature AIDA-I\*. As shown in Fig. 1C, the 11kDa fragment appears after renaturation of the protein, but is absent from an uncleaved mutant, used as control (see below). This experiment shows that AIDA-I is processed by an autocatalytic event. We noted, however, that this *in vitro* cleavage was less efficient than the fast and complete cleavage that is usually observed in bacteria expressing AIDA-I.

Indeed, uncleaved protein remains even after 6 days of renaturation. This could be due to slow and incomplete folding, as shown below.

We used mass spectrometry on *in vitro* refolded mature AIDA-I\* in order to determine the precise location of the cleavage observed. As expected, the same C-terminal peptide as for native AIDA-I, N<sub>837</sub>IILNPTKES<sub>846</sub>, was found, confirming that the cleavage occurs at the usual site during the *in vitro* refolding, between S846 and A847 (19). To exclude the possibility that cleavage resulted from a contaminating protease, we performed refolding experiments with a variant of AIDA-I\* bearing a point mutation in one putative catalytic residue (mutant detailed below) and showed that this protein is not cleaved in the same conditions (Fig.1C). Taken together, those results are the first direct proof of an auto-catalytic cleavage activity for the AIDA-I autotransporter.

**Correlation between folding and cleavage.** Appearance of mature AIDA-I\* is strikingly slow, taking more than one day. Protein folding is generally thought to be faster, with folding rates calculated in seconds. However, extremely slow folding was already observed with the native pertactin (40), which possesses the same  $\beta$ -helical structure that is predicted for AIDA-I (39). We next wanted to gain insight in the folding kinetics of AIDA-I in order to assess if cleavage could correlate with protein folding. Evidence of the *in vitro* folding of AIDA-I\* was obtained by monitoring the changes in the far UV circular dichroism (CD) signal of the protein. As shown in Fig.1D, gain of structure was evident in the CD signal over time after exchange with renaturation buffer but the apparition of secondary structure was remarkably slow. The time course monitored at 218 nm (Fig.1E) can be fitted to single exponential decay with rate constant of  $0.74 \pm 0.17 \text{ min}^{-1}$ , which is consistent with the kinetic of appearance of mature AIDA-I.

To confirm this observation, we used a common characteristic of folded protein: resistance to proteolysis. At different time points during refolding, a sample was submitted to a 30 minute protease treatment. Most of mature AIDA-I\* is resistant to the protease at all time points where it is visible, while the proprotein was completely susceptible (Fig.2). Taken together, these results strongly suggest that mature AIDA-I\* is indeed folded, and that folding and cleavage of AIDA-I\* are two closely related events.

**The pH-dependence of the AIDA-I *in vitro* folding.** The efficiency of AIDA-I\* *in vitro* folding was examined at different pH values between 4 and 10. We observed that the



refolding buffer used had a marked effect on the refolding kinetics of AIDA-I\*. At pH10, the protein remains unfolded, even at 145h after buffer exchange (Fig.3). Consequently, no cleavage was apparent at all time points and the protein was completely susceptible to protease digestion, confirming the absence of renaturation at pH10.

Conversely, at pH6 the renaturation of AIDA-I\* was closed to what observed at pH8, with similar folding rate, cleavage kinetic and protease resistance (Fig.3).

Unexpectedly, at pH4, the protein seems almost completely folded at 2h after buffer exchange (Fig.3). However, the protein refolded at pH4 remains uncleaved, even after 6 days of renaturation. We try to restore the cleavage using a pH shift from pH4 to pH8. However, even at 145h after the pH shift, no cleavage was observed. We also submitted a sample to a 30 minute protease treatment to verify the proper folding. Surprisingly, the protein was partially susceptible to protease, leaving a stable fragment of approximately 23 kDa. These results suggest that the renaturation of the AIDA-I\* protein at pH4 is fast but aberrant, observation that can explain the absence of proteolytic cleavage. All together, these results highlight the pH-dependence of the AIDA-I\* *in vitro* folding and reinforce the observation that cleavage is closely related to the proper folding of the protein.

**Mapping of the region involved in AIDA-I cleavage.** To support our model of autocatalytic cleavage, we decided to map the region of the protein involved in the cleavage reaction. To address this question, we used different deletion mutants and assessed the capacity of the resulting proteins to process themselves. The AIDA $\Delta$ N (deletion of amino acids 45 to 225), AIDA $\Delta$ C2 (deletion of amino acids 224 to 667), AIDA $\Delta$ SS (deletion of the entire signal sequence, amino acids 2 to 49) and the AIDA-I\* mutant were tested for their ability to process themselves (Fig.4A). All mutants were co-expressed with the glycosyltransferase Aah and whole-cell lysates were probed with antibodies against the His tag. As shown in Fig.4B, all deletion mutants were cleaved, with apparition of mature proteins at the expected sizes. For the AIDA $\Delta$ C2 protein, 2 different bands appeared, one at the expected size and another one slightly lower that may correspond to a degradation product. We have previously observed that this mutant is indeed unstable (37). Overall, these results suggest that the AIDA-I catalytic site is located between A667 and T953, a region encompassing the conserved junction domain of autotransporters (amino acids T851 to L952) as well as a part of mature AIDA-I with the cleaved peptidic bond between S846 and A847).

**Two acidic residues are required for AIDA-I cleavage.** None of the classical sequences for aspartyl, serine or metalloproteases could be identified in AIDA-I. Furthermore, as this protein lacks cysteine residues, we ruled out the possibility of cysteine protease activity. Other autotransporters of *E. coli* belonging to the SPATE family are auto-catalytically cleaved by a novel mechanism involving an asparagine and an aspartate residue embedded in the  $\beta$ -barrel (29). Also, some bacterial aspartate proteases do not possess the classical D(S/T)G consensus motif, as for instance is the case with omptins (41). We therefore hypothesized that aspartate residues may be responsible for AIDA-I cleavage.

To test this hypothesis we exchanged several aspartate residues between A667 and T953 with alanines (Fig.5A). This amino acid was chosen because it is less likely to negatively influence the catalytic activity and the folding of the mutated proteins. For all mutants, whole-cell lysates were obtained and probed with antibodies against the His tag. As shown in Fig.5B, most of the mutations did not affect the biogenesis or the cleavage of the protein. For the D878A mutant, however, only the proprotein form is observed, suggesting an inhibition of cleavage. The fact that processing is completely abolished strongly suggests that this aspartate plays a catalytic role. To test the requirement for the carboxyl group of the aspartate in position 878, as would be expected if it were to be catalytic, we changed this amino acid by an asparagine or by a glutamate. Conservation of the carboxyl group maintained the cleavage, whereas the modification by an asparagine residue abolished completely the processing of AIDA-I.

As acidic proteases usually form catalytic dyad or triad, we decided to change into alanine residues a few amino acids near D878. We chose E845, N849, E871, N874, R879 and E897. The cleavage was slightly altered in two mutants, E845A and N849A, which are localized near the cleavage site. One mutation, however, E897A, resulted in a complete abolition of processing. The cleavage was still inhibited when the glutamate was changed into glutamine, highlighting a role for this second carboxyl group in the cleavage reaction. We also modified all aspartate residues (D1021, D1078, D1117, D1170, D1175, D1233, D1234) predicted to be located in the extracellular loop of the  $\beta$ -barrel of AIDAc and which, therefore, could potentially be spatially close to the cleavage site but none of the mutations resulted in a reduction of the processing (data not shown). Taken together, our results suggest that D878 and E897 form a catalytic dyad of carboxylic amino acids, which is common in acidic proteases.

**Characterization of the D878N uncleaved mutant.** We compared the adhesive, self-aggregation and biofilm formation activities mediated by the uncleaved D878N mutant and the WT protein. We observed that the D878N mutant is as functional as the wild type in all assays (data not shown). This is consistent with our previous observation obtained with a protein mutated at the cleavage site (15). The diminished AIDA-I cleavage in this mutant did not affect either the functionality of the protein.

We purified the WT and the D878N proteins from the outer membrane in their native conformations and compared their secondary structure, stability and relative protease resistance. The Far UV CD spectrum for the D878N mutant is slightly different from the spectrum of the WT protein (Fig.6A). However, by following the thermal denaturation at 218 nm continuously while changing the temperature between 25°C and 80°C, we observed that the thermal denaturation of the two proteins is almost identical (Fig.6B). Lastly, the conformation of the uncleaved mutant was assayed by protease resistance, and, as illustrated in Fig.6C, the two proteins showed the same resistance to trypsin, in agreement with the thermal denaturation. We noted that it is only AIDA<sub>c</sub> that seems to be sensitive to the protease but did not investigate further. Overall, these results show that the uncleaved mutant behaved almost indistinguishably from the wild-type protein.

***In vitro* refolding of the D878N uncleaved mutant.** We next wanted to compare the *in vitro* refolding kinetics of the D878N mutant with that of the wild-type protein. We introduced the D878N mutation in the plasmid pAngH $\Delta$ SS $\Delta$  $\alpha\beta$ , generating the D878N\* protein. We purified D878N\* from inclusion bodies and performed the *in vitro* renaturation as described above with the wild-type protein. As shown in Fig.7A, in contrast to the WT protein, the D878N\* mutant remained in the uncleaved form, even at 145 hours after the buffer exchange. The absence of cleavage was confirmed by the lack of the 11 kDa peptide that was present with the wild-type AIDA-I\* protein (Fig.1C).

We also monitored the changes in the far UV CD of D878N\*. As for the WT protein, the folding of the uncleaved mutant was especially slow (Fig.7B). We noted a small difference in the final CD spectrum for the D878N mutant in comparison to the WT protein, as observed for the native proteins purified from the outer membrane (Fig.6A). Again, this suggests the existence of a subtle difference in secondary structure between the two proteins. The refolding time course monitored at 218 nm could be fitted to single

exponential decay with rate constant of  $1.09 \pm 0.39 \text{ min}^{-1}$ , similar to the kinetics of the wild-type protein (Fig.7C). To confirm the renaturation of the D878N mutant, we submitted the protein to a 30 minute protease treatment. As shown in Fig. 7D, the percentage of protein resistant to protease increased with time, confirming the *in vitro* refolding of this protein. Taken together, these results suggest that the *in vitro* folding kinetics of the AIDA-I\* and the D878N\* proteins are similar, arguing that *in vitro* cleavage observed with wild-type AIDA-I\* is directly linked to the presence of the amino acid D878.

**AIDA-I is cleaved by an intramolecular mechanism.** In a previous publication, we have shown that the cleavage of AIDA-I was independent of its expression level (15). This observation was consistent with an intramolecular processing but a proof that this is indeed the case is still missing. We first tried the *in vitro* refolding at pH8 with different concentration of AIDA-I\* between 150 and 400  $\mu\text{g/ml}$ , in order to evaluate the effect of protein concentration on the kinetic of cleavage. All time courses for mature AIDA-I\* appearance were fitted to single exponential association with rate constant that are not significantly different ( $0.77 \pm 0.22 \text{ min}^{-1}$  for 150  $\mu\text{g/ml}$ ;  $0.53 \pm 0.19 \text{ min}^{-1}$  for 200  $\mu\text{g/ml}$ ;  $0.70 \pm 0.26 \text{ min}^{-1}$  for 250  $\mu\text{g/ml}$  and  $0.52 \pm 0.18 \text{ min}^{-1}$  for 400  $\mu\text{g/ml}$ ), which suggest an intramolecular reaction. With the mutant D878N we can also directly address this question. We purified the WT and D878N proteins in their native conformation and incubated the two proteins together. As shown in Fig.8A, the level of D878N proprotein remains unchanged, even after 2 hours of incubation with the WT protein, showing that no intermolecular cleavage could be observed. A similar experiment was performed *in vivo* with bacteria expressing the WT or D878N proteins. After co-incubation of the two bacterial cultures for 30 minutes, no intermolecular cleavage was observed (Fig.8B). Lastly, we expressed both proteins on the surface of the same bacteria. As shown in Fig.8C, the D878N mutant remains uncleaved in the presence of a functional wild-type protein. Taken together, these results showed that autoproteolysis of AIDA-I is intramolecular.

## Discussion

In this study, we prove the auto-catalytic nature of the cleavage of the AIDA-I autotransporter by following the *in vitro* refolding of a mutant lacking its C-terminal  $\alpha$ -helix and  $\beta$ -barrel. We noted that *in vitro* cleavage was less efficient than processing *in vivo*. This could be due to incomplete folding. Indeed, we observed that folding and processing

occur at similar rates, which strongly suggests that the two events are closely related. This observation is also an evidence arguing against the cleavage resulting from of a contaminant protease, since the folding of AIDA-I is strikingly slow and an exogenous protease would most likely cleave AIDA-I faster. Furthermore, we observed that the *in vitro* refolding of AIDA-I is dramatically pH-dependant and without folding (pH10) or with an aberrant folding (pH4), the protein remains uncleaved, arguing again for a strong association between folding and cleavage.

We identified by deletion analysis a 286 amino acids region directly involved in the processing of AIDA-I. This region includes the junction domain, which ends just before the cleavage site, and the C-terminal part of the AIDA-I functional domain. We chose to mutagenize acidic residues of this region because of the observation that *in vitro* refolding and processing of AIDA-I occurred even in presence of serine and cysteine protease inhibitors. Some acidic proteases, like the omptin family, rely on catalytic aspartate residues that cannot be identified by classical consensus motifs (42). We found that D878 and E897 in the junction region are involved in the processing of AIDA-I. The complete elimination of cleavage when these residues are mutated strongly suggests that they play a catalytic role. Mutations that preserve the acidic side chain were tolerated, giving more evidence that the carboxylic acid groups are directly involved in catalysis. The requirement of D878 for catalytic activity was confirmed by *in vitro* refolding experiments with the D878N mutant, showing that it can refold but remains uncleaved. We observed that the uncleaved mutant presents a normal global conformation, as this mutant is as functional and as stable as the wild-type AIDA-I. Lastly, we confirmed that the cleavage of AIDA-I relies on an intramolecular mechanism of processing.

As shown in Fig.9A, our results are consistent with a three-dimensional structural model of the junction region and the AIDA-I cleavage site (the amino acids between I807 and L956), which we generated using the known crystal structure of *B. pertussis* pertactin (39). This model predicts that D878 and E897 are located close to the cleaved peptidic bond, which is found in a loop. This gives an explanation as to why folding and cleavage occur at similar rates: the junction domain of AIDA-I and the region encompassing the cleavage site have to be properly folded for the correct positioning of the catalytic residues in relation to the cleaved peptidic bond. The model also explains why mutations such as E845A or N849A lead to a decrease in the processing efficiency, since these changes would likely destabilize the loop bearing the bond that has to be cleaved.

The mechanism of autocatalytic cleavage we describe here for AIDA-I is unique among the monomeric autotransporter family. Alignment of the junction regions sequences of several autotransporters highlights that D878 of AIDA-I is conserved among autotransporters, whether they are processed or not (Fig.9B). However, E897 is unique to AIDA-I among auto-catalytic cleaved autotransporters, which could explain why this mechanism of proteolysis is unique to AIDA-I. Our model also differs from the two previously described intramolecular mechanisms of cleavage for autotransporters: the processing mediated by endogenous serine protease domains or SPATE and pertactin-like autotransporters.

The cleavage mechanism for several other autotransporters, including the Ag43 aggregation factor, remains completely unknown. As Ag43 and AIDA-I are closely related in sequence and possess similar functional features (18), it is tempting to hypothesize that a similar mechanism of processing involving carboxylic amino acids exists for Ag43. However, Ag43 is cleaved at a different site compared to the one of AIDA-I (43), and in a region that has only poor homologies with known autotransporter structures.

The question of the function of AIDA-I cleavage remains. In autotransporters that are cleaved, the processing can play different roles, such as the release of cytotoxic domain in the extracellular milieu (44), keeping a protein localized at a pole (45), or controlling the expression level of an adhesin (46). The intramolecular nature of AIDA-I processing excludes the possibility that it serves to regulate the protein level at the surface of the bacteria. In addition, we did not observe that cleavage is important for function, as uncleaved mutants resulting from mutations of the catalytic residues or mutations of the residues of the cleaved peptidic bond are as functional as the wild-type protein. This puzzling observation is at least consistent with the fact that we could not find any conditions that lead to the release of mature AIDA-I in the extracellular milieu (15). We can hypothesize that in the context of an animal infection, a specific condition may cause the release of mature AIDA-I, which would be beneficial for the bacteria, for example, to escape the immune response. Alternatively, cleavage could generate a new unique surface representing an interaction site with other unknown proteins or host factors, which could play a role during infection. This mechanism was recently observed with the EscU protein, a structural component of the inner membrane ring of the type three secretion system in *E. coli* (47). This last hypothesis would be consistent with the small differences we observed

between CD spectra for the D878N mutant or the WT protein, in their native conformation or after *in vitro* folding.

Self-processing of autotransporters is a frequent feature and occurs by a variety of different mechanisms, with our study providing yet another one. This diversity is quite surprising and strongly suggests that this processing plays an important role. Further investigation is therefore required to establish what this role might be in the case of AIDA-I.

### Acknowledgments

We thank François Lépine for performing the mass spectrometry experiments. This work was supported by financial contributions from the Groupe de Recherche et d'Études sur les Maladies Infectieuses du Porc (GREMIP), The Canadian Institutes of Health Research (CIHR, 84578), the Canada Research Chair program and the Canada Foundation for Innovation (201414), and a graduate fellowship to M.-E.C. from the Natural Sciences and Engineering Research Council of Canada (NSERC, 85297596).

### References

1. Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., and Ala'Aldeen, D. (2004) *Microbiol. Mol. Biol. Rev.* **68**(4), 692-744
2. Henderson, I. R., and Nataro, J. P. (2001) *Infect. Immun.* **69**(3), 1231-1243
3. Girard, V., and Mourez, M. (2006) *Res. Microbiol.* **157**(5), 407-416
4. Peterson, J. H., Szabady, R. L., and Bernstein, H. D. (2006) *J. Biol. Chem.* **281**(14), 9038-9048
5. Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J., and Gros, P. (2004) *EMBO J.* **23**(6), 1257-1266
6. Barnard, T. J., Dautin, N., Lukacik, P., Bernstein, H. D., and Buchanan, S. K. (2007) *Nat. Struct. Mol. Biol.* **14**(12), 1214-1220
7. Oliver, D. C., Huang, G., Nodel, E., Pleasance, S., and Fernandez, R. C. (2003) *Mol. Microbiol.* **47**(5), 1367-1383
8. Renn, J. P., and Clark, P. L. (2008) *Biopolymers* **89**(5), 420-427
9. Mogensen, J. E., Tapadar, D., Schmidt, M. A., and Otzen, D. E. (2005) *Biochemistry* **44**(11), 4533-4545
10. Benz, I., and Schmidt, M. A. (1989) *Infect. Immun.* **57**(5), 1506-1511

11. Niewerth, U., Frey, A., Voss, T., Le Bouguenec, C., Baljer, G., Franke, S., and Schmidt, M. A. (2001) *Clin. Diag. Lab. Immunol.* **8**(1), 143-149
12. Chapman, T. A., Wu, X. Y., Barchia, I., Bettelheim, K. A., Driesen, S., Trott, D., Wilson, M., and Chin, J. J. (2006) *Appl. Environr. Microbiol.* **72**(7), 4782-4795
13. Zhang, W., Zhao, M., Ruesch, L., Omot, A., and Francis, D. (2007) *Vet. Microbiol.* **123**(1-3), 145-152
14. Sherlock, O., Schembri, M. A., Reisner, A., and Klemm, P. (2004) *J. Bacteriol.* **186**(23), 8058-8065
15. Charbonneau, M. E., Berthiaume, F., and Mourez, M. (2006) *J. Bacteriol.* **188**(24), 8504-8512
16. Benz, I., and Schmidt, M. A. (2001) *Mol. Microbiol.* **40**(6), 1403-1413
17. Charbonneau, M. E., Girard, V., Nikolakakis, A., Campos, M., Berthiaume, F., Dumas, F., Lepine, F., and Mourez, M. (2007) *J. bacteriol.* **189**(24), 8880-8889
18. Klemm, P., Vejborg, R. M., and Sherlock, O. (2006) *Int. J. Med. Microbiol.* **296**(4-5), 187-195
19. Suhr, M., Benz, I., and Schmidt, M. A. (1996) *Mol. Microbiol.* **22**(1), 31-42
20. Benz, I., and Schmidt, M. A. (1992) *Mol. Microbiol.* **6**(11), 1539-1546
21. Shere, K. D., Sallustio, S., Manassis, A., D'Aversa, T. G., and Goldberg, M. B. (1997) *Mol. Microbiol.* **25**(3), 451-462
22. van Ulsen, P., van Alphen, L., ten Hove, J., Fransen, F., van der Ley, P., and Tommassen, J. (2003) *Mol. Microbiol.* **50**(3), 1017-1030
23. van Ulsen, P., Adler, B., Fassler, P., Gilbert, M., van Schilfgaarde, M., van der Ley, P., van Alphen, L., and Tommassen, J. (2006) *Microbes Infect.* **8**(8), 2088-2097
24. Turner, D. P., Marietou, A. G., Johnston, L., Ho, K. K., Rogers, A. J., Wooldridge, K. G., and Ala'Aldeen, D. A. (2006) *Infect. Immun.* **74**(5), 2957-2964
25. Fink, D. L., Cope, L. D., Hansen, E. J., and Geme, J. W., 3rd. (2001) *J. Biol. Chem.* **276**(42), 39492-39500
26. Coutte, L., Willery, E., Antoine, R., Drobecq, H., Loch, C., and Jacob-Dubuisson, F. (2003) *Mol. Microbiol.* **49**(2), 529-539
27. Serruto, D., Adu-Bobie, J., Scarselli, M., Veggi, D., Pizza, M., Rappuoli, R., and Arico, B. (2003) *Mol. Microbiol.* **48**(2), 323-334
28. Pohlner, J., Halter, R., Beyreuther, K., and Meyer, T. F. (1987) *Nature* **325**(6103), 458-462



29. Dautin, N., Barnard, T. J., Anderson, D. E., and Bernstein, H. D. (2007) *EMBO J.* **26**(7), 1942-1952
30. Henderson, I. R., and Owen, P. (1999) *J. Bacteriol.* **181**(7), 2132-2141
31. Nguyen, V. Q., Caprioli, R. M., and Cover, T. L. (2001) *Infect. Immun.* **69**(1), 543-546
32. Vandahl, B. B., Pedersen, A. S., Gevaert, K., Holm, A., Vandekerckhove, J., Christiansen, G., and Birkelund, S. (2002) *BMC Microbiol.* **2**, 36
33. Ohnishi, Y., Beppu, T., and Horinouchi, S. (1997) *J. of Biochem.* **121**(5), 902-913
34. Litwin, C. M., and Johnson, J. M. (2005) *Infect. Immun.* **73**(7), 4205-4213
35. Litwin, C. M., Rawlins, M. L., and Swenson, E. M. (2007) *Infect. Immun.* **75**(11), 5255-5263
36. Charbonneau, M. E., and Mourez, M. (2008) *Res. Microbiol.* **159**(7-8), 537-544
37. Charbonneau, M. E., and Mourez, M. (2007) *J. Bacteriol.* **189**(24), 9020-9029
38. Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., and Kelley, L. A. (2008) *Proteins* **70**(3), 611-625
39. Emsley, P., Charles, I. G., Fairweather, N. F., and Isaacs, N. W. (1996) *Nature* **381**(6577), 90-92
40. Junker, M., Schuster, C. C., McDonnell, A. V., Sorg, K. A., Finn, M. C., Berger, B., and Clark, P. L. (2006) *Proc. Natl. Acad. Sci. USA* **103**(13), 4918-4923
41. Vandeputte-Rutten, L., Kramer, R. A., Kroon, J., Dekker, N., Egmond, M. R., and Gros, P. (2001) *EMBO J.* **20**(18), 5033-5039
42. Hritonenko, V., and Stathopoulos, C. (2007) *Mol. Membr. Biol.* **24**(5-6), 395-406
43. Klemm, P., Hjerrild, L., Gjermansen, M., and Schembri, M. A. (2004) *Mol. Microbiol.* **51**(1), 283-296
44. Navarro-Garcia, F., Canizalez-Roman, A., Luna, J., Sears, C., and Nataro, J. P. (2001) *Infect. Immun.* **69**(2), 1053-1060
45. d'Hauteville, H., Dufourcq Lagelouse, R., Nato, F., and Sansonetti, P. J. (1996) *Infect. Immun.* **64**(2), 511-517
46. Fink, D. L., and St Geme, J. W., 3rd. (2003) *J. Bacteriol.* **185**(5), 1608-1615
47. Zarivach, R., Deng, W., Vuckovic, M., Felise, H. B., Nguyen, H. V., Miller, S. I., Finlay, B. B., and Strynadka, N. C. (2008) *Nature* **453**(7191), 124-127
48. Otto, B. R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J. G., Mizutani, K., Park, S. Y., and Tame, J. R. (2005) *J. Biol. Chem.* **280**(17), 17339-17345

## Footnotes

The abbreviations used are: Aah: Autotransporter Adhesin Heptosyltransferase; AIDA-I: Adhesin Involved in Diffuse Adherence; CD: circular dichroism; GnHcl: guanidinium hydrochloride; HisG: protein tag consisting of six histidine residues and a glycine; IPTG: isopropyl- $\beta$ -thiogalactopyranoside; LB: Luria-Bertani medium; PBS: Phosphate Buffered Saline; PMSF: phenylmethanesulphonyl fluoride; SAATs: Self-Associating Autotransporters; SPATE: Serine Protease Autotransporters of Enterobacteriaceae; TBS: Tris-Buffered Saline.

## Figures Legends

**Fig.1: Cleavage of AIDA-I\* *in vitro*.** *A.* Schematic diagram of the AIDA-I\* construct, showing plasmid pAngH (WT) and pAngH $\Delta$ SS $\Delta$  $\alpha\beta$  ( $\Delta$ SS $\Delta$  $\alpha\beta$  or AIDA-I\*). The filled box represents the HisG tag at the N terminus of mature AIDA-I. *B.* Inclusion bodies containing the AIDA-I\* protein were collected, solubilized in 0.1 M Tris-Hcl pH8 and 6 M GnHcl and purified by immobilized metal affinity chromatography. Renaturation of the AIDA-I protein was performed by first exchanging the buffer on a 1-ml His trap HP column for 2 hours with renaturation buffer at pH8 (TBS, 0.1 % Triton X-100), and the protein was then left at 4°C. At different times after the buffer exchange, aliquots were taken and boiled in SDS-PAGE loading buffer. The proteins were then resolved by SDS-PAGE and the gel stained with Coomassie blue. The uncleaved (proprotein, circles) and cleaved (mature AIDA-I\*, arrowheads) forms of the protein are indicated. The intensities of the bands were quantified and normalized to the maximal intensity. *C.* The WT and the uncleaved (D878N\*) proteins at 145h after buffer exchange were separated by SDS-PAGE on a 16% Tris-Tricine gel. The star indicates the C-terminal cleaved peptide. *D.* Far-UV CD spectra of the AIDA-I\* protein at different times after buffer exchange. The ellipticities were recorded between 205 and 260 nm. *E.* The ellipticities at 218 nm were recorded at different times after the initial renaturation. MRE, mean residual ellipticity.

**Fig.2: Protease sensitivity of AIDA-I\* during refolding.** Limited proteolysis of pure AIDA-I\* conducted for 30 minutes with 37  $\mu\text{g.ml}^{-1}$  (final concentration) of trypsin at different times after buffer exchange. The proteins were resolved by SDS-PAGE and the gel stained with Coomassie blue.

**Fig.3: Effect of pH on the *in vitro* folding of AIDA-I\*.** Renaturation of the AIDA-I\* protein was performed by first exchanging the buffer on a 1-ml His trap HP column for 2 hours with renaturation buffer at pH8 (2 mM Tris pH8, 150 mM NaCl, 0.1 % Triton X-100). An equal amount of protein was diluted in buffer with different pH values (50 mM sodium citrate pH4 or pH6; 50 mM Tris pH8; 50 mM sodium carbonate pH10; all supplemented with 150 mM NaCl and 0.1 % Triton X-100) and left at 4°C. *A.* Far-UV CD spectra of the AIDA-I\* protein at different times following buffer exchange for pH4 (top) pH6 (middle) or pH10 (bottom). The ellipticities were recorded between 205 and 260 nm. MRE, mean residual ellipticity. *C.* At different times after the buffer exchange, aliquots were taken and boiled in SDS-PAGE loading buffer. *D.* Limited proteolysis at 145h after buffer exchange performed as described in Fig.2. For pH4, the proteolysis was performed after a pH shift at pH8. The proteins were resolved by SDS-PAGE and the gel stained with Coomassie blue. The uncleaved (proprotein, circles) and cleaved (mature AIDA-I\*, arrowheads) forms of the protein are indicated.

**Fig.4: Mapping of the region involved in AIDA-I cleavage.** *A.* Schematic diagram of the deletion constructs, showing plasmid pAngH (WT), pAngH $\Delta$ N ( $\Delta$ N), pAngH $\Delta$ C2 ( $\Delta$ C2), pAngH $\Delta$ SS ( $\Delta$ SS) and pAngH $\Delta$ SS $\Delta$  $\alpha\beta$  ( $\Delta$ SS $\Delta$  $\alpha\beta$  or AIDA-I\*). The filled box represents the HisG tag at the N terminus of mature AIDA-I. *B.* Whole-cell lysates were obtained from overnight cultures of C600 harboring an empty vector (-), expressing wild-type AIDA-I (WT), or expressing the  $\Delta$ N,  $\Delta$ C2,  $\Delta$ SS or  $\Delta$ SS $\Delta$  $\alpha\beta$  mutants were obtained and probed with anti-His antibodies, which allowed detection of the proprotein (circles) and mature AIDA-I (arrowheads). All plasmids were transformed in bacteria expressing the Aah glycosyltransferase.

**Fig.5: Mutational analysis of the region involved in AIDA-I cleavage.** *A.* Predicted secondary structure of the region of AIDA-I encompassing the amino acids between N708 and S953; E:  $\beta$ -strand; x: no prediction. The positions corresponding to modified amino acids are indicated. The dashed line indicates the cleavage site. The numbering of amino acids corresponds to the position in full-length pre-proprotein. *B.* Whole-cell lysates were obtained from overnight cultures of C600 harboring an empty vector (-), plasmid pAgH (WT), or one of the plasmid bearing a point mutation in pAgH. Proteins separated by SDS-

PAGE were probed by immunoblotting with antibodies against the His tag, which allowed the detection of the proprotein (Pro., circles) and the mature AIDA-I (arrowheads).

**Fig.6: Characterization of the D878N protein.** *A.* Far-UV CD spectra of the purified AIDA-I wild-type protein (grey circles) and the uncleaved D878N mutant (black squares). The ellipticities were recorded between 205 and 260 nm. *B.* The ellipticities at 218 nm of the WT protein (grey circles) or the D878N mutant (black squares) were recorded with the temperature varying between 25°C and 80°C at a rate of 5°C per min. MRE, mean residual ellipticity. *C.* Purified uncleaved (D878N) and cleaved (WT) proteins were incubated in the presence of different concentration of trypsin. The proteins were resolved by SDS-PAGE and the gel stained with Coomassie blue.

**Fig.7: Renaturation and folding kinetic of D878N\* mutant.** The D878N\* protein was purified as describe in Fig.1. *A.* At different times after buffer exchange, aliquots were taken and boiled in SDS-PAGE loading buffer. The proteins were then resolved by SDS-PAGE and the gel stained with Coomassie blue. The uncleaved (proprotein D878N\*, circles) form of the protein is indicated. *B.* Far-UV CD spectra of the D878N\* protein. The ellipticities were recorded between 205 and 260 nm. The CD spectrum of the WT AIDA-I\* after 145h of renaturation is indicated. *C.* The ellipticities at 218 nm were recorded at different time after the initial renaturation. MRE, mean residual ellipticity. *D.* Limited proteolysis performed as described in Fig.2. The proteins were resolved by SDS-PAGE and the gel stained with Coomassie blue.

**Fig.8: Intramolecular cleavage of AIDA-I.** *A.* Wild-type AIDA-I (WT) or the uncleaved (D878N) proteins were incubated separately or together for 30 minutes at room temperature, resolved by SDS-PAGE, and the gel stained with Coomassie blue. *B.* Whole-cells lysates were obtained from overnight cultures of *E. coli* C600 harboring plasmids pTRC99a (-), pAgH (WT) or pAgHD878N (D878N) co-incubated for 30 minutes at 30°C. The samples were resolved by SDS-PAGE and probed with anti-His tag antibodies, which allowed the detection of the proprotein (circle) and the mature AIDA-I (arrowheads). *C.* Whole-cells lysates from overnight cultures of *E. coli* C600 harboring plasmids pTRC99a and/or PACYC184 (-), pAgH, pACYC-AgH, pAgHD878N or pACYC-

AgH and pAgHD878N were resolved by SDS-PAGE and probed with anti-His tag antibodies.

**Fig.9: Structural model and alignment of the cleavage site of AIDA-I.** *A.* A three-dimensional structural model of the junction region and the cleavage site of AIDA-I (amino acids I807-L956). The amino acids constituting the cleavage site, S846 and A847, are highlighted in cyan, and the amino acids D878 and E897 are highlighted in red and depicted with their side chains. The bottom panel shows a close-up view of the cleavage site and the D878 and E897 residues using the same color scheme. *B.* Alignment of the junction region of several autotransporters. The secondary structures features, either modeled (for AIDA-I) or experimentally determined (for Hbp (48) and Prn (39)) are indicated. The positions of S846, A847, D878 and E897 are highlighted in cyan and red, as in *A.* The accession numbers for the sequence data and the amino acids range used for the alignment are: ABS20376, amino acids 807-956 (*E. coli* AIDA-I); Q9XD84, amino acids 470-629 (*E. coli* TibA); P39180, amino acids 540-704 (*E. coli* Ag43); O88093, amino acids 919-1073 (*E. coli* Hbp); Q7BS42, amino acids 915-1058 (*E. coli* Pic); O68900, amino acids 836-991 (*E. coli* Pet); AAC44731, amino acids 851-1022 (*E. coli* EspC); Q7BSW5, amino acids 840-996 (*E. coli* EspP); CAA88252, amino acids 911-1062 (*S. flexneri* SepA); CAC05837, amino acids 572-739 (*S. flexneri* IcsA); Q03035, amino acids 416-569 (*B. pertussis* Prn); AAA51646, amino acids 551-706 (*B. pertussis* BrkA); CAA45708, amino acids 828-984 (*H. influenzae* IgA protease); and P45387, amino acids 800-977 (*H. influenzae* Hap).

**Table I. Primers used in this study**

Primer	Sequence
SS $\Delta\alpha\beta$ .....	5'-GGTATTTAACCAGTTAACTTCCCACATCTGATACCCGG-3'
D712A.....	5'-GGACTAACATTCAGAACGCCGGCAATTCATTCTCAATCT TGC-3'
D724A.....	5'-GCTGAAAACATGCTTTTGAACAGAGCTCTCAGGTAGTGG-3'
D772A.....	5'-GCAGTAGTCAATGCAGCCATGGCTGTCAACCAGAATGCC-3'
D785A.....	5'-GCCTATATAAACATTAGTGCACAGGCAACAATTAATGG-3'
E845A.....	5'CTTAATCCTACAAAAGCTAGCGCAGGTAATACTCTTACCG-3'
N849A.....	5'CCTACAAAAGAAAGTGCCGGCGCTACTCTTACCGTGTC-3'
E871A.....	5'CTCTGGTGGTGTGCTTGCCGGCGATAATCACTTACGG-3'
D873A.....	5'-CTCTGGTGGTGTCTAGAAAGGAGCTAATCACTTACGG-3'
N874A.....	5'-GCTTGAAGGAGATGCTAGCCTTACGGACCGTCTGGTGG-3'
D878A.....	5'-GGAGATAATCACTTACGGCCCGACTAGTGGTGAAAGGTAAT ACC-3'
D878E.....	5'-GGAGATAATCACTTACGGAACGACTAGTGGTGAAAGGTAAT ACC-3'
D878N.....	5'-GGAGATAATCACTTACGAACCGGTTGGTGGTGAAAGGTAAT ACC-3'
R879A.....	5'-GGAGATAATCACTTACGGACGCACTAGTGGTGAAAGGTAA TACC-3'
E897A.....	5'-GGTCAAAGTGACATCGTTTATGTTAACGCAGATGGCAGTGGT GG-3'
E897Q.....	5'-GGTCAAAGTGACATCGTTTATGTTAACCAAGATGGCAGTGGT GG-3'
D906A.....	5'-GGTGGTCAGACTCGAGCTGGTATTAATATTATTCTGTAGAG GG-3'
D933A.....	5'-CGCGTAGTTGCCGGAGCATATGCTTACACACTGCAGAAAG G-3'
D1021A.....	5'-GGTAAGCTTAATGCCGGCCAAAATAAAACAACAACC-3'
D1117A.....	5'-GCATCAGTGAAAGGTGCCGGCCTGGAAGAAG-3'
D1170-5A.....	5'-GGGGTTACGCCGGCTACACATCAGGAGGCTAACGGAACG-3'
D1078A.....	5'-GCTGCCAGAAACACGCTAGCTGGTTATTCTGTCTGGGG-3'
D1233-4A.....	5'-GGTGTAAAATGAGTGCTGCTAGCCAGTTGTTGTCAGG-3'

Figure 1

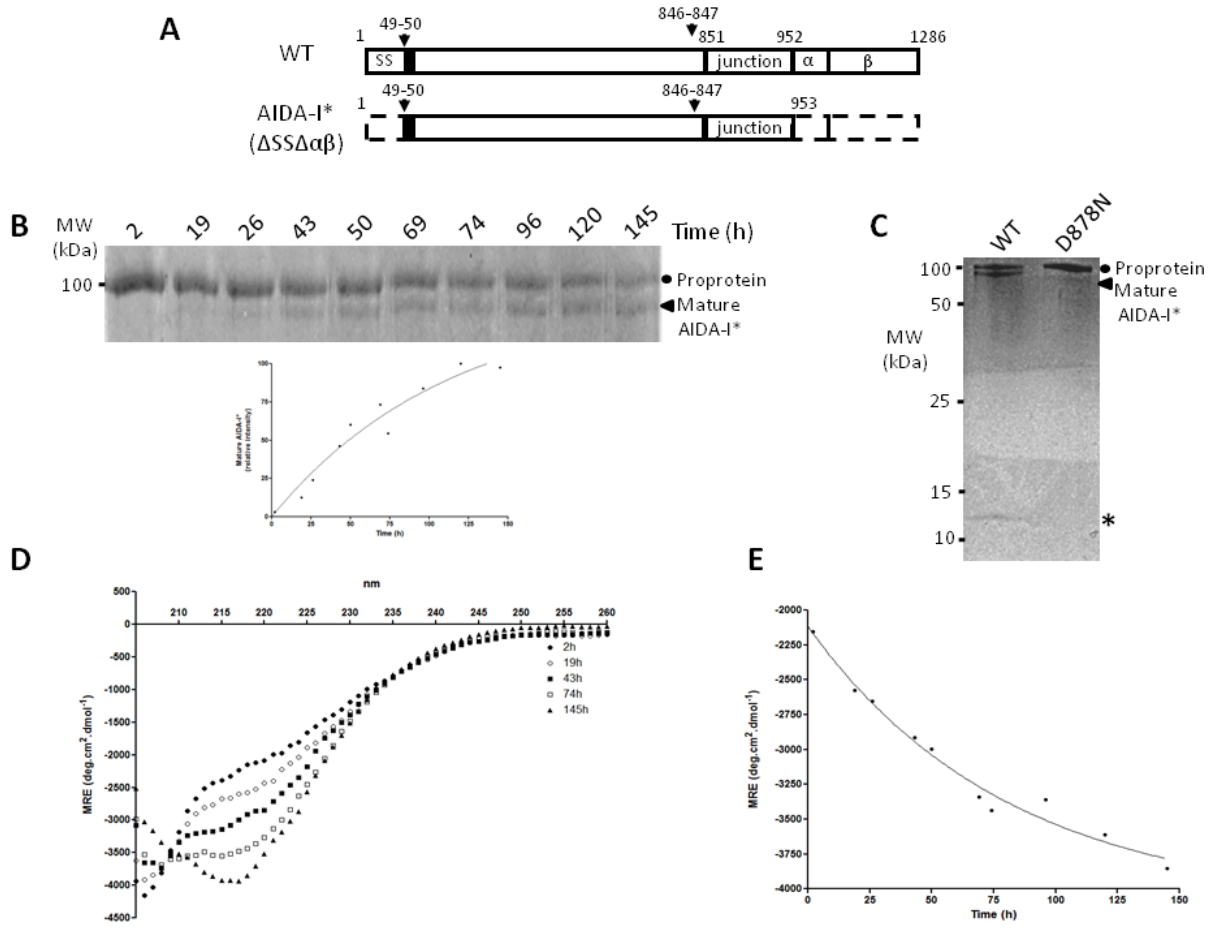






Figure 3

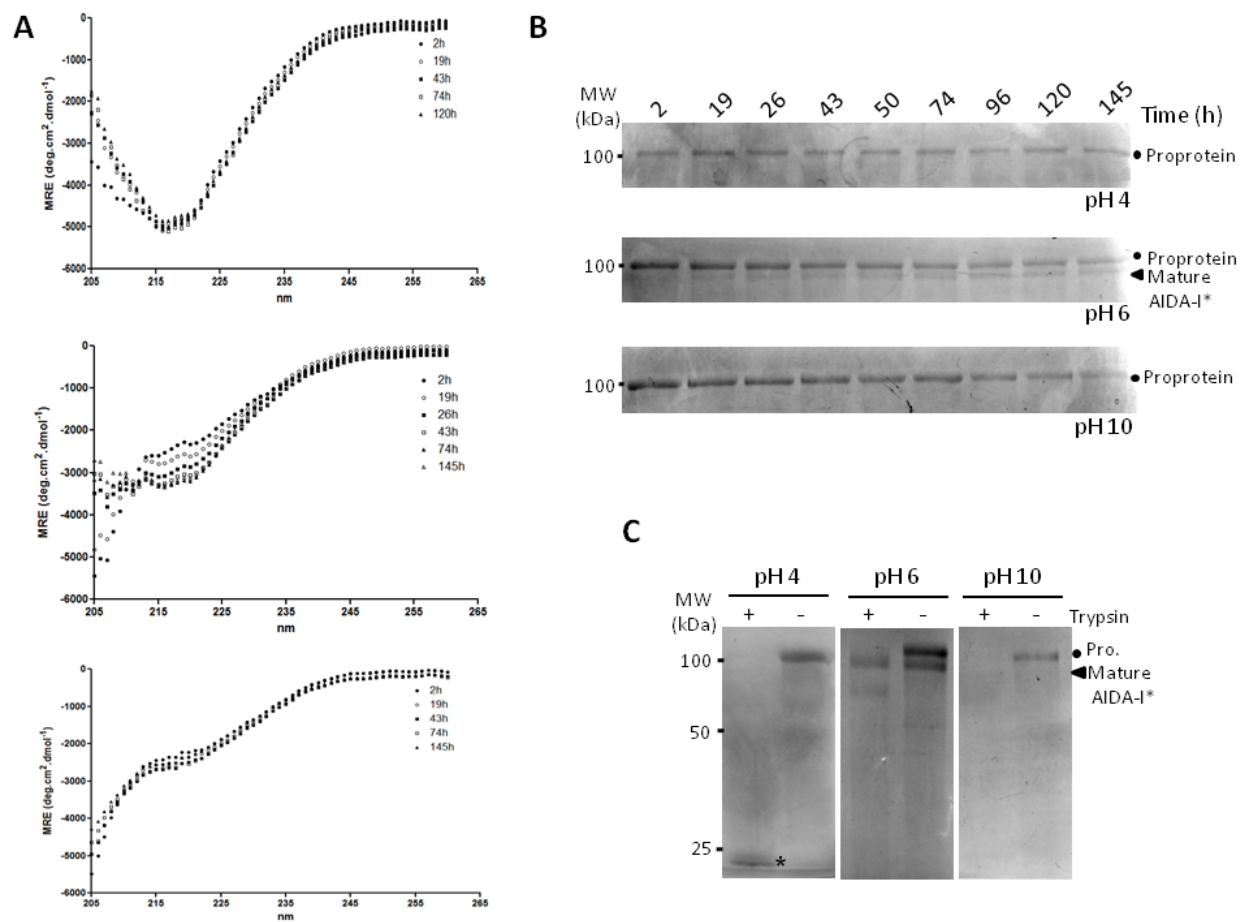


Figure 4

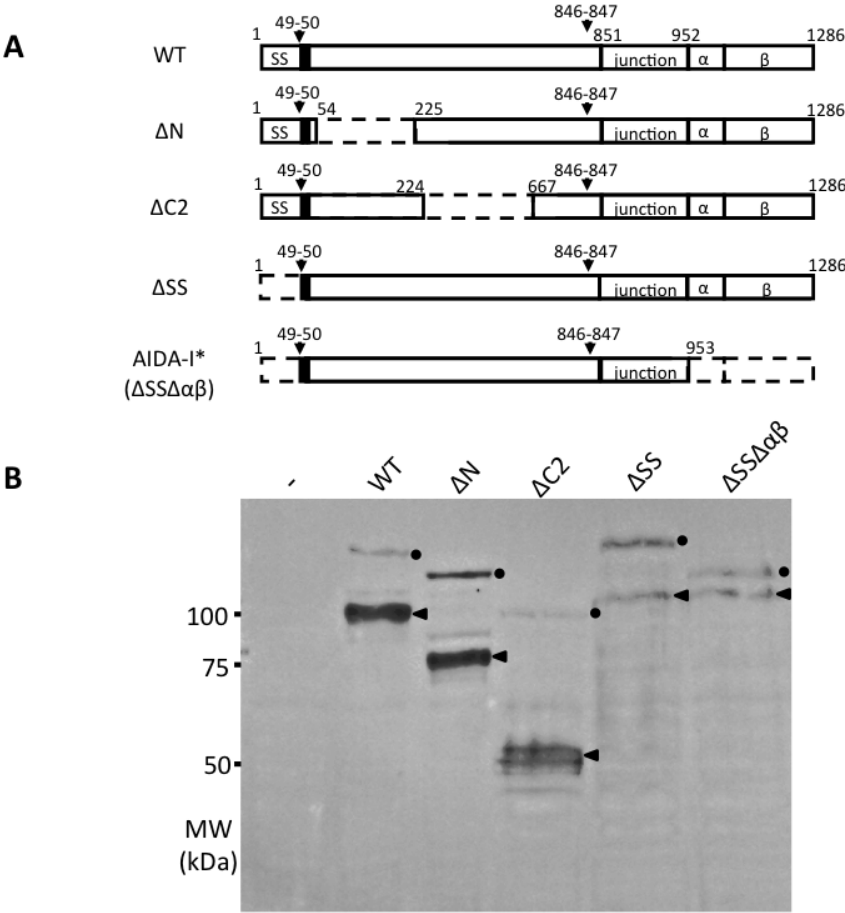


Figure 5

A

↓712                      ↓724  
NIQNDGNFILNLAENYDFETELSGSGVLVKDNTGIMTYAGTLTQAQGVNVKN  
EEE EEEEE EEE EEEE EEEEE EEEEE EEE  
↓772                      ↓785  
GGIIFDSAVVNADMAVNQAYINISDQATINGSVNNNGSIVINNSIINGNITNDADLSFG  
EEEEEE EEE EEEEE EEEEE EEEEE EEEEE EEE E  
845 ↓                      ↓849                      871 ↓                      ↓879  
TAKLLSATVNGSLVNNKNIILNPTKESAGNLTVSNYTGTGPSVISLGGVLEGDNSLTDR  
EEEE EEEEE EEEEE EEEEE EEEEE EEEEE EEEEE EEEEE EEEEE  
↓897                      ↓906                      ↓933  
LVVKGNTSGQSDIVYVNEGSGGQTRDGINIISVEGNSDAEFLKNRVVAGAYDYTLQKG  
EEE EEEEEEEEEEE xxx EEEEE EEEEE EEEEE EEEEE  
NESGTDNKGWYLT  
EEE EEEEEEE

B

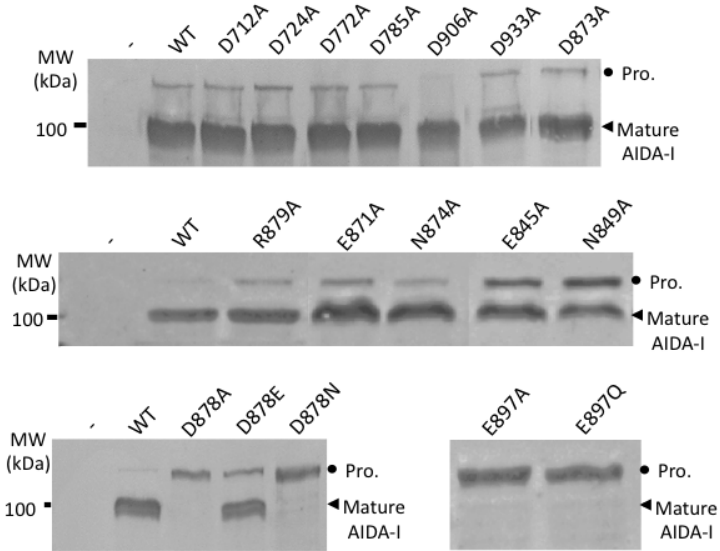


Figure 6

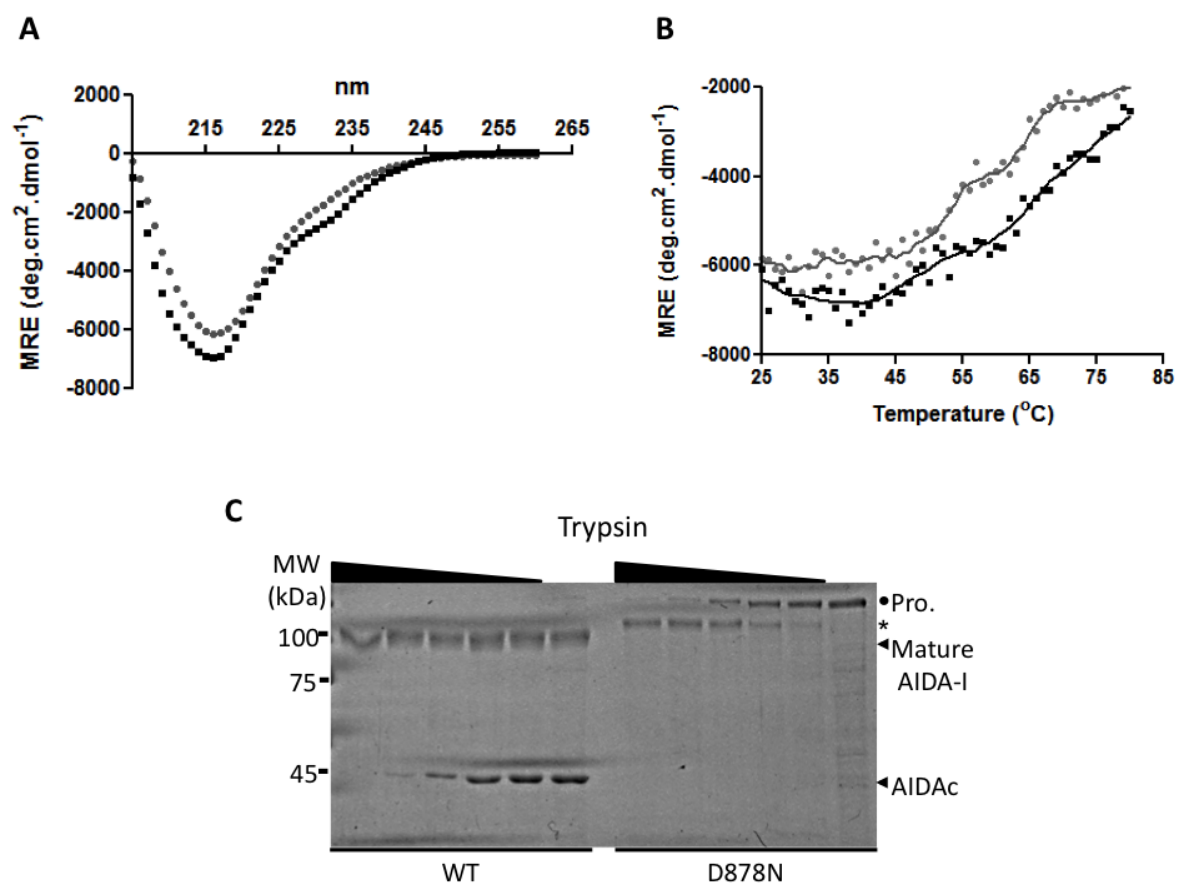


Figure 7

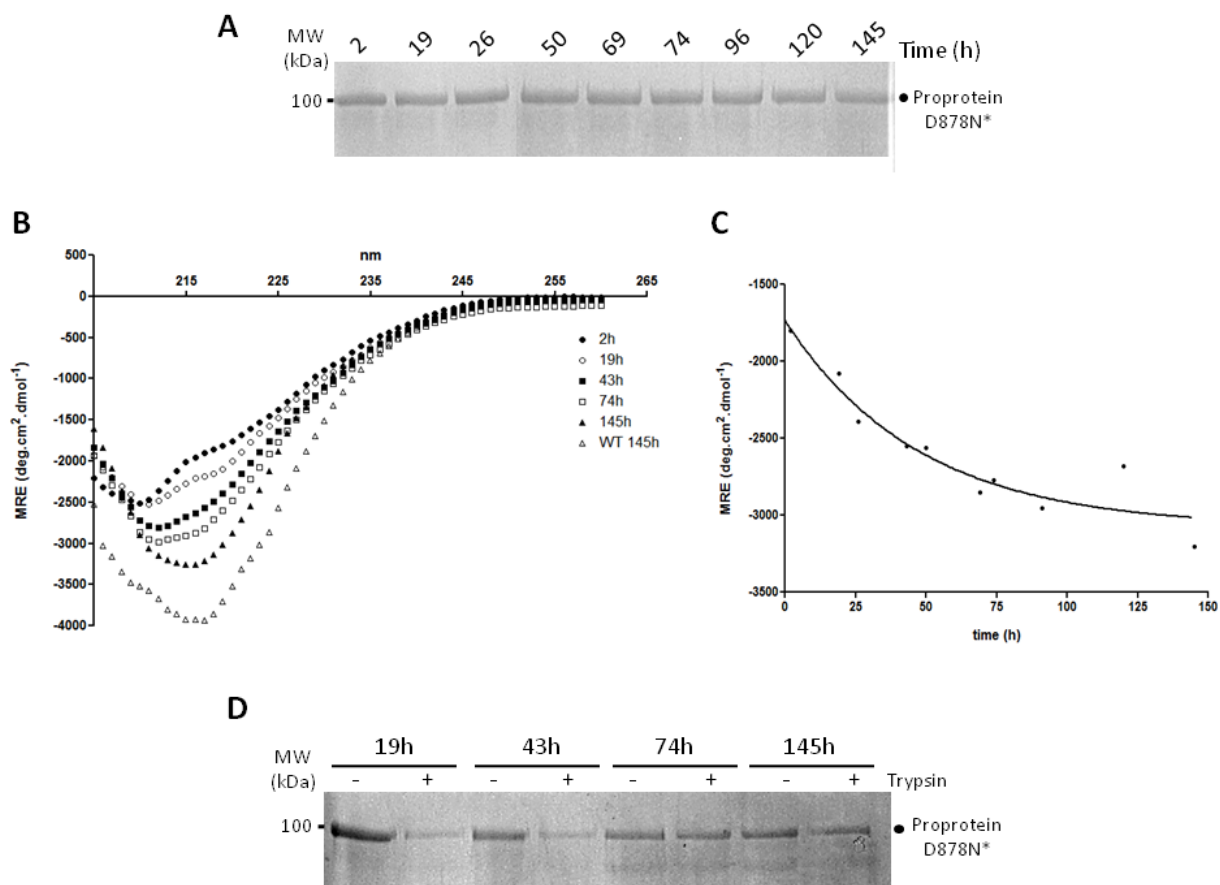


Figure 8

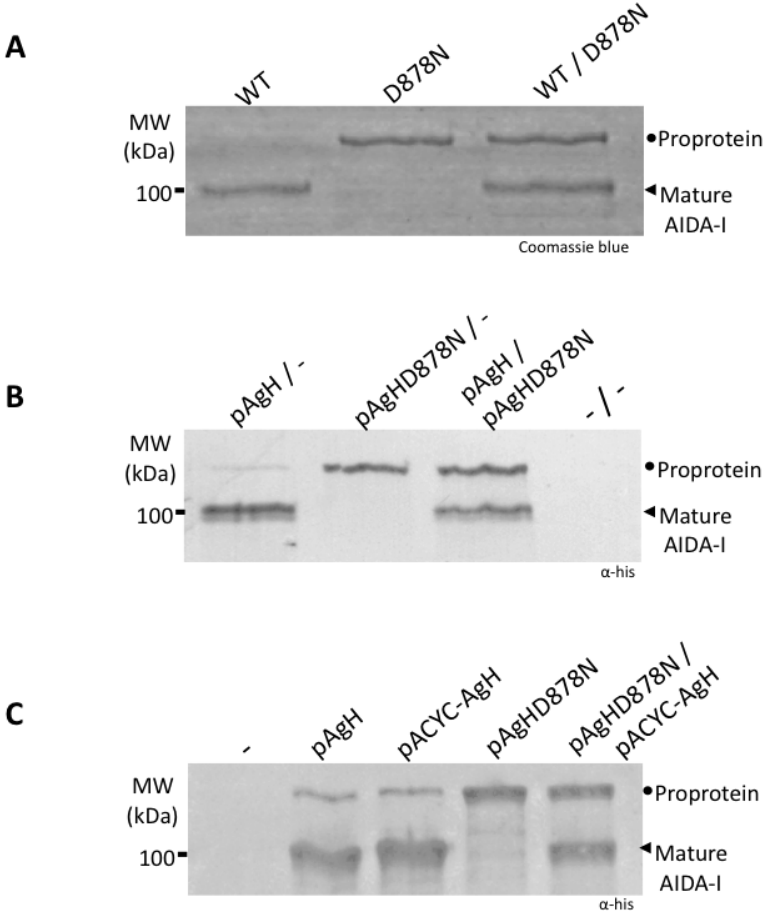
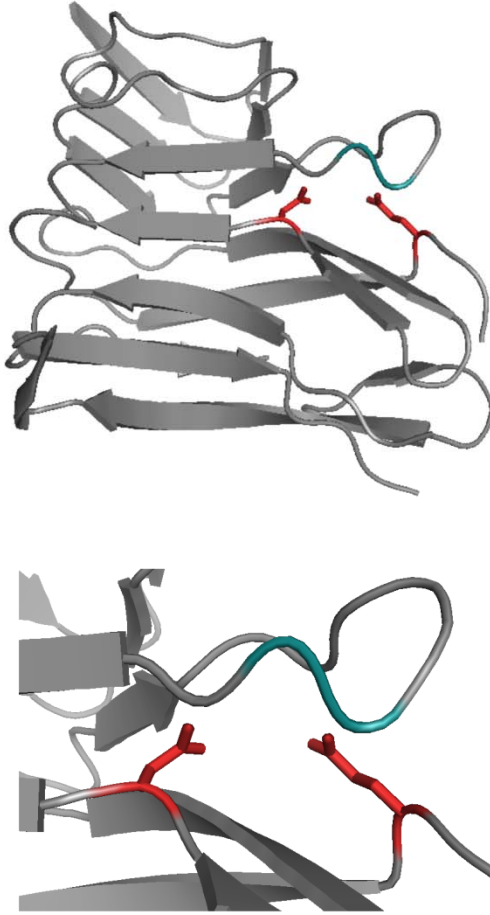


Figure 9



AIDA-I INGSVNNNGSIVINN-----SIINGNITNDADLSFGTAKLLSATVNGSLVNNKNIILNPTK-----  
 TibA LDNGTILTGNVAGDSTGAGDMAVK-GASVHLDGDSVTGAL---TLDNGTVDFRPSSTTRMTF---  
 Ag43 LTGSTVLNGAIDPTN-----VTLASGATWNIIPDNATVQSVVDDLHSHAGQIHFTSTRTG-----  
 Hbp NGYRNWSGSLNAPDA-----TVSMTDTQWSMNGNSTAGNM---KLNRTIVGFNGG-----

Pic EGYNAAFNGAITGGRA-----DVMHNLWTLGGDSAIHSL---TVRNSRISSEGD-----  
 Pet DGFDTSYQSSIKAAQS-----SLAMNNALWKVTGNSSELKKL---NSTGSMVLFNGG-----  
 EspC RGYNAVLSGNTGEOQS-----TVNMNNALWYSDGNSITIGTL---KSTGGRVELGGG-----  
 SepP GGFVDSYRGSVNPASA-----SATMNNTWQLTGDSEALKTL---KSTNSMVYFSDSAN-----  
 SepA KGYNAVLRGSI TGAQS-----TVNMNIALWYSDGKSEAGAL---KAKGSRIELGDGK-----  
 IcsA KGATLNFSGMNQTVNTLLNSGTVLNNINAPFLPDPVIVTGNMTEKNGHVLNNSG-----  
 Prn LASQARWTGATRAVDS-----LSIDNATWVMNDNSVNGAL---RLASDGSVDFQPAE-----

BrkA VASGAAWHGATQVLQS-----ATLGKGGTWWVNADSRVQDM---SMRGRGRVFOAPAP-----  
 IgA VLGKANLFTIQSRGN---SQVRLTENSHWHLTGNSDVHQL---DLANGHIHLNSADNSNN-----  
 Hap SLKNSHFHQIQGDKG---TTVLENAWTMPDSTTLQNL---TLNNSITILNSAYSASSNNTPR-----

AIDA-I -----ESAGNLTIVSNYTGTPGFSVISLGGVLEGDNSLTLRLVVK-GNTSGOSDIVVYK  
 TibA -----AFQAVSLALGSLSGS-GTFQMNNDI---ASHTGDMLNVA-GNASGNFVLDIKNT  
 Ag43 -----KFVPATLKVKNLNGQNGTISLVRFPD-MAQNNADRIVDDGRATGKTLINLVNA  
 Hbp -----TSPFETLITDNLDAVQSAFVVRITDL-----NKADKLVIN-KSATGHNSIWNF

Pic -----RTFRITLVNKL DATGSDPFLRITDL-----KNADKINVT-EKATGSDNSLNVSF  
 Pet -----KNI FNTLTVDEL TTSNSAFVVRITN-----TQADQLIVK-NKLEGANLLVDF  
 EspC -----KDFATLRVKELNANNATFLMHTN-----TSQADQLNVT-NKLLGSNNITLVDF  
 SepP -----NKKFHTLVDELATSNAYAMRITNL-----SESDKLEVK-KHLSGNNITLVDF  
 SepA -----HFATLQVKEL SADNTFLMHTN-----NSRADQLNVT-DKLSGNSNITLVDF  
 IcsA -----SNVGQTYVQKGNWHGKGGILSLGAVLGNDSKTDRIETA-GHSGITVYVAVITNE  
 Prn -----AGRFRVLMVDTLAGS-GLFRMNVFA---DLGLSDKLVVM-RDASQOHLWVRNS

BrkA -----EASYKTLTLOTLDG-NGVFLNNTNV---AAGONDQLRVT-GRADGQHRVLVRNA  
 IgA -----VTKYNTLTVNSLSG-NGSFYYLITDL---SNKQGDVVVT-KSATGNFTLQVADK  
 Hap RRSLETETPTSAEHRFNLTIVNGKLSGGQTFQFTSSLF---GYKSDIKKLS-NDAGDYILSVRNT

AIDA-I DGSG---GQTRDGINIISVEGNSDAEFSL-----KNRVVAGAYDYTLQKNGESCTDNKGNVYLTSHL  
 TibA GLELV---SAGAPLQVVTGGGDAAPTLL-----KGGKVDACTWYGL-----SKENNTWYLKADT  
 Ag43 GNSASGLATSCKGIQVVEANGATTEEGA--FVQGNRLQAGAFNYSL-----NRDSDSEWYLRSEN  
 Hbp LKKE---SNKDTLIDIPVSAPEATADNL-----FRASTRVVGFSVDVTPILSVRKEGDKKEWVLDGQY

Pic MNNE---AQGOALNIPLVTAAGTSAEM-----FACRTRVTGFSRVTPTLHVDTSGGNPKWVLDGPK  
 Pet IEKK-GNDKNGLNIDLKAPENTSQDV-----FKTETQTI GFSVDVTP EIKQEKDGKSVWTLTGK  
 EspC LNKP---ASEMNVLTITAFKGSDEKT-----FTAGTQQIGFSNVTPVISTEKTDATKMWLTGYQ  
 SepP LQKP---TPEKQLNIELVSAFKDITNENV-----FKASKQTI GFSVDVTPVITRETDKTIWLSLTGYN  
 SepA LNKP---ASEMSVLTITAFKGSDEKT-----FTAGTQQIGFSNVTPVISTEKTDATKMWLTGYQ  
 IcsA GGSG---DKTLEGVQIISTDSSDKNAFI-----QKGRIVAGSYDTRKQGTASGLNNTNKWLYTSQM  
 Prn GSEP---ASANTMLLVOTPRGSAATFTLA---NKDGKVDIGTYRYRLA-----ANGNGQSLVGAKE

BrkA GGEA---DSRGARLGLVHTQGGQGNATFRLANVGKA--VDLGTWRYSLA-----EDPKTHVWSLQKQAG  
 IgA TGER---NHNELTLDASKAQRDHLNVS--LVGNTVDLGAWKYKL-----RNVNGRYDLYNPE  
 Hap GKPE---ETLEQLTVESKDNQPLSDKLFKFTLENDHVDAGALRYKLV-----KNDGEPRLHNP I

## ARTICLE 6

**A structural motif is the recognition site for a new family of bacterial O-glycosyltransferase**

Molecular Microbiology, 2012, Mar., Volume 83 (5), pages 894-907

## Contribution de l'auteur pour cet article

En tant que premier auteur de cet article, j'ai effectué la plupart des expérimentations décrites, à l'exception des mutants chromosomiques chez *Citrobacter rodentium*, les clonages de l'homologue de Aah de *C. rodentium* et de l'homologue de AIDA-I dans *Cronobacter sakasaki*, ainsi que les expériences requérant la spectrométrie de masse. J'ai écrit la totalité de l'article.

## Contribution scientifique de l'article

Le mécanisme de reconnaissance du substrat par les O-glycosyltransférases bactériennes est inconnu pour la plupart des enzymes étudiées. Dans cet article nous démontrons que ce n'est pas la séquence primaire en acides aminés qui est reconnue par la glycosyltransférase Aah, mais plutôt un motif structural. De plus, nous avons identifié de nombreux homologues de Aah, dont un chez *Citrobacter rodentium* qui possède un mode d'action similaire à Aah. À partir de ces observations, nous suggérons que Aah est le prototype pour une nouvelle famille de O-glycosyltransférases bactériennes.

Cet article a fait l'objet d'un micro-commentaire (Otzen, D., N for AsN - O for StrOcture? A strand-loop-strand motif for prokaryotic O-glycosylation, Molecular Microbiology, 2012, Mar., Volume 83 (5), pages 879-83).



**A structural motif is the recognition site for a new family of bacterial protein O-glycosyltransferases**

Marie-Ève Charbonneau<sup>1</sup>, Jean-Philippe Côté<sup>1</sup>, M. Florencia Haurat<sup>2</sup>, Bela Reiz<sup>3</sup>, Sébastien Crépin<sup>4</sup>, Frédéric Berthiaume<sup>1</sup>, Charles M. Dozois<sup>4</sup>, Mario F. Feldman<sup>2</sup> and Michael Mourez<sup>1\*</sup>

Running title: Substrate specificity of the Aah glycosyltransferase

<sup>1</sup>Canada Research Chair on Bacterial Animal Diseases, Université de Montréal, Faculté de Médecine Vétérinaire, St-Hyacinthe, 3200 Sicotte, St-Hyacinthe, Québec, J2S 7C6, Canada. <sup>2</sup>Department of Biological Sciences and <sup>3</sup>Mass spectrometry facility, Department of chemistry, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada. <sup>4</sup>INRS-Institut Armand-Frappier, 531 Boul. des Prairies, Laval, Québec, H7V 1B7, Canada.

Keywords: Protein glycosylation, heptosyltransferase, AIDA-I, autotransporter, *Escherichia coli*

**Abstract**

The *Escherichia coli* Adhesin Involved in Diffuse Adherence (AIDA-I) is a multifunctional protein that belongs to the family of monomeric autotransporters. This adhesin can be glycosylated by the AIDA-associated heptosyltransferase (Aah). Glycosylation appears to be restricted to the extracellular domain of AIDA-I, which comprises imperfect repeats of a 19 amino acid consensus sequence and is predicted to form a  $\beta$ -helix. Here, we show that Aah homologs can be found in many Gram-negative bacteria, including *Citrobacter rodentium*. We demonstrated that an AIDA-like protein is glycosylated in this species by the Aah homolog. We then investigated the substrate recognition mechanism of the *E. coli* Aah heptosyltransferase. We found that a peptide corresponding to one repeat of the 19 amino acid consensus is sufficient for recognition and glycosylation by Aah. Mutagenesis studies suggested that, unexpectedly, Aah recognizes a structural motif typical of  $\beta$ -helices, but not a specific sequence. In agreement with this finding, we observed that the extracellular domain of the *Bordetella pertussis* pertactin, a  $\beta$ -helical polypeptide lacking the 19 amino acid consensus sequence, could be glycosylated by Aah. Overall, our findings suggest that Aah represents the prototype of a new large family of bacterial protein *O*-glycosyltransferases that modify various substrates recognized through a structural motif.

## Introduction

Protein glycosylation has been extensively studied in eukaryotes. The identification in the past few years of numerous glycosylated prokaryotic proteins indicates that this modification is in fact distributed among all kingdoms of life (Nothaft & Szymanski, 2010). Protein glycosylation provides diverse functions, such as modulating protein stability and resistance against proteases, affecting intermolecular interactions and more. Well-studied prokaryotic glycoproteins are often extracellular and include the subunits of S-layers, flagella, type IV pili, and many adhesins, such as autotransporters (Schmidt *et al.*, 2003, Szymanski & Wren, 2005). Recently, a new family of bacterial oligosaccharyltransferases, represented by the HMW1C protein of *Haemophilus influenzae*, has been described and seems to be involved in glycosylation of surface proteins with adhesive function (Grass *et al.*, 2010). In addition, other reports describing general *N*- and *O*-glycosylation systems from *Campylobacter*, *Neisseria* and *Bacteroides* ssp., identified another class of bacterial glycoproteins that are required inside the cell rather than at the cell surface to carry out their functions (Szymanski *et al.*, 1999, Vik *et al.*, 2009, Fletcher *et al.*, 2009). Furthermore, hundreds of *B. fragilis* proteins from all cellular locations, including essential proteins, are predicted to be glycosylated (Fletcher *et al.*, 2011). Thus, it is clear that glycosylation is an important and widespread bacterial protein modification.

How glycosyltransferases recognize their protein substrates is intensely investigated. In *C. jejuni*, it has been suggested that the primary consensus motif D/E-X-N\*-X-S/T is recognized in proteins that are N-glycosylated, where X is any amino acid except proline (Kowarik *et al.*, 2006). It should be noted, however, that variations of this glycosylation motif have been described, such as the PglB oligosaccharyltransferase of *Desulfovibrio desulfuricans* (Ielmini & Feldman, 2011). A similar consensus also exists for eukaryotic protein N-glycosylation (Yan & Lennarz, 1999). For *O*-glycosylation, however, the recognized sequence is less clearly defined beyond the modified serine or threonine. In *Neisseria meningitidis* proteins, *O*-glycosylation is found in low complexity regions rich in serine, threonine, proline and asparagine, and glycans are often found in the motif S\*-A-P-A (Vik *et al.*, 2009). However, this motif is not sufficient or required for glycosylation. Similarly, *O*-glycosylation is associated with proline-rich regions of proteins of *Clostridium thermocellum* or *Mycobacterium tuberculosis*. However, no motif sufficient or required for glycosylation was identified (Gerwig *et al.*, 1993, Dobos *et al.*, 1996). The motif D-S/T\*-A/I/L/V/T/M was proposed as a consensus sequence recognized by protein *O*-

glycosyltransferases in *Bacteroides fragilis*, and the N-S\* and N-T\*-T motifs were found in glycoproteins of *Elizabethkinga meningoseptica* (Fletcher et al., 2009, Plummer et al., 1995). It is clear from all these studies that no general consensus sequence for bacterial O-glycosylation can be highlighted. This suggests that specific oligosaccharyltransferases may possess distinct substrate specificities or that other signals are recognized besides primary structure. More information about substrate recognition by bacterial O-glycosyltransferases is therefore needed.

One of the few glycoproteins found in *Escherichia coli* is the Adhesin Involved in Diffuse Adherence (AIDA-I) (Benz & Schmidt, 2001). This plasmid-encoded protein was first identified in a strain isolated from a case of infantile diarrhea (Benz & Schmidt, 1989). Since then, this adhesin has been associated with pathogenic strains of *E. coli* involved in neonatal and post-weaning diarrhea in piglets (Chapman et al., 2006, Niewerth et al., 2001, Zhang et al., 2007). Besides its role in adhesion, AIDA-I has also been shown to mediate biofilm formation, bacterial auto-aggregation, and invasion of epithelial cells (Sherlock et al., 2004, Charbonneau et al., 2006). This protein belongs to the monomeric autotransporter family of the type V secretion pathway (Henderson et al., 2004). Like all autotransporters, AIDA-I comprises an N-terminal signal sequence, a C-terminal membrane-embedded domain and a central functional extracellular domain. With the auto-aggregation factor antigen 43 (Ag43) and the TibA adhesin-invasin protein, AIDA-I forms a family of autotransporters called Self-Associating Autotransporters (SAATs) (Klemm et al., 2006). These three proteins share the same virulence attributes and possess imperfect repetitions of an identical 19 amino acid consensus in the extracellular domain. These repeated sequences are predicted to fold into a right-handed  $\beta$ -helix, a structural motif often found in the extracellular domain of autotransporters (Kajava & Steven, 2006).

The AIDA-associated heptosyltransferase (Aah) is responsible for the glycosylation of AIDA-I and diverts heptose residues from the lipopolysaccharide biosynthesis pathway (Benz & Schmidt, 2001). The *aah* gene is located immediately upstream and forms an operon with the *aidA* gene. Similarly, the TibA autotransporter is glycosylated by the TibC specific heptosyltransferase and the genes coding for these proteins are organized like *aah-aidA* (Lindenthal & Elsinghorst, 1999). These two glycosyltransferases are functionally interchangeable (Moormann et al., 2002). Ag43 is also found in a glycosylated form in wild-type uropathogenic *E. coli* even though a specific glycosyltransferase has not been

identified to date (Sherlock et al., 2006). Thus, heptosylation seems to be shared among these proteins.

AIDA-I can be modified by addition of up to 19 heptose molecules, and these sugars are added on serine and threonine residues, confirming the *O*-linked type of glycosylation (Charbonneau *et al.*, 2007). An alignment of the sequences of AIDA-I or Ag43 peptides glycosylated by Aah suggested a potential recognition motif for Aah (Knudsen *et al.*, 2008). However, the motif corresponds mainly to the repeated sequence of 19 amino acids and therefore whether the motif is necessary or sufficient remains to be tested.

In this study, we observed that homologs of Aah are found in many bacterial species, suggesting that Aah is a paradigm for a new large family of bacterial protein *O*-glycosyltransferases. These putative glycosyltransferases were sometimes associated with autotransporters or large putative adhesins, many of which did not possess the 19 amino acid repeated consensus. This suggested that the motif recognized by these glycosyltransferases is different from the consensus sequence found in AIDA-I. We therefore investigated the substrate specificity of Aah. Our results unexpectedly suggest that Aah recognizes a structural motif typical of the right-handed  $\beta$ -helix, and not a specific amino acid sequence. To the best of our knowledge, that a bacterial glycosyltransferase relies on structural cues has not been previously documented. A better understanding of the *O*-glycosylation processes in bacteria may uncover new avenues for the exploitation of these pathways for biotechnological glycoengineering purposes.

## Results

**A new family of bacterial protein glycosyltransferases.** A BLAST search with the Aah amino acid sequence (CAC43407) retrieved 44 bacterial proteins with highly significant E-values (E-value  $\leq 10^{-73}$ ). The *aah* homologs coding for these proteins are found in 3 branches of *Proteobacteria*: *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* (Fig. 1). Four bacterial strains that belong to the *Alphaproteobacteria* branch (*Bradyrhizobium* sp. ORS278, *G. bethesdensis* CGDNIH1, *A. pasteurianus* IFO 3283-01 and *G. oxydans* 621H) possess more than one copy of an Aah homolog. All Aah homologs possess the heptosyl/glycosyltransferase family 9 pfam01075 domain and sequence alignment shows that these proteins indeed share a high level of

sequence similarity (Fig. S1). These results suggest that Aah is the paradigm for a family of glycosyltransferases distributed widely among proteobacterial classes.

By examining the genomic regions immediately flanking the *aah* homologs, we often identified putative monomeric autotransporters characterized by the conserved domain pfam03797. Furthermore, visual examination of the sequence of these autotransporters often revealed the 19 amino acid consensus characteristic of AIDA-I (Table 1 and see below). This organization, similar to that of *aah* and *aidA*, suggested that the Aah homologs could modify an associated putative autotransporter. To test this hypothesis, we first used antibodies that detect specifically the heptoses grafted by Aah onto AIDA-I. Indeed, we previously showed that antibodies that were directed against the glycosylated extracellular domain of AIDA-I, do not efficiently recognize the non-glycosylated form of AIDA-I and can be used as an indicator of glycosylation by Aah (Charbonneau et al., 2007). Moreover, these AIDA-I antibodies also react specifically with glycosylated TibA (Côté, J-P. and Mourez, M., unpublished observation). We selected three bacterial strains from two different classes of *Proteobacteria* possessing an Aah homolog and a flanking putative autotransporter that has the 19 amino acid consensus sequence. As shown in Fig. 2A, when transferred to a membrane, whole cell lysates of *Citrobacter rodentium* DBS100, of *Cronobacter sakazakii* HBP 2178, and to a lesser extent of *Burkholderia vietnamiensis* G4, reacted with our AIDA-I antibodies. *E. coli* strain 2787 and its isogenic *aidA* mutant were used as control and show the specificity of our antibodies. Thus, this result suggests that a glycosylation similar to that of Aah-mediated glycosylation of AIDA-I occurs naturally in *C. rodentium*, *C. sakazakii* and *B. vietnamiensis*.

To confirm that an Aah homolog indeed modified an associated putative autotransporter, we used the *C. rodentium* strain DBS100. This strain possesses an Aah homolog (YP\_003368456; renamed Aah<sub>Cr</sub>) and an associated putative autotransporter (YP\_003368457; renamed AIDA<sub>Cr</sub>). Chromosomal deletions of the *aah*<sub>Cr</sub> gene or the *aidA*<sub>Cr</sub> gene were constructed and tested for appearance of a glycosylated protein. As shown in Fig. 2B, a band of approximately 230 kDa was revealed by probing whole cell lysates of *C. rodentium* with our antibodies directed against AIDA-I, in agreement with the result obtained with the dot-blot. The molecular weight observed is higher than the 160 kDa predicted for AIDA<sub>Cr</sub>. This slower electrophoretic mobility suggested that the protein is modified, in agreement with the fact that the band was revealed by our glycosylation-specific antiserum. The band disappeared in mutant strains lacking *aah*<sub>Cr</sub> or *aidA*<sub>Cr</sub> and in

the *aah<sub>Cr</sub>* mutant the band reappeared by complementation with a plasmid expressing the *aah<sub>Cr</sub>* gene under the control of an IPTG-inducible promoter (Fig. 2B). These experiments confirm that the immunoreactive band is AIDA<sub>Cr</sub> modified by Aah<sub>Cr</sub>.

To further prove that the Aah homologs identified here recognize similar protein substrates as Aah or TibC, we used a more distant bacterium, *C. sakazakii*. As shown in Fig. 2A, at least one protein of *C. sakazakii* is glycosylated. Flanking the Aah homolog, this bacterium possesses a gene coding for a putative monomeric autotransporter (ESA\_02084 and renamed here *aidA<sub>Cs</sub>*), which encompasses multiple repeats of the same 19 amino acid consensus (Fig. 3A). We decided to test the glycosylation of AIDA<sub>Cs</sub> by Aah when recombinantly co-expressed in *E. coli* K-12. As shown in Fig. 3B, in the presence of a plasmid expressing AIDA<sub>Cs</sub> (predicted molecular weight of 234 kDa), a protein of approximately 250 kDa is visible in whole cell extract stained with Coomassie blue. A second band is also present at a lower molecular weight. This second band could be a specific processing product of AIDA<sub>Cs</sub>, similar to the autoprocessing observed with AIDA-I and Ag43. In the presence of Aah, these bands were also revealed by our anti-AIDA-I antibody and by the specific ProQ Emerald 300 glycan detection kit, suggesting that AIDA<sub>Cs</sub> is indeed glycosylated. In summary, our results suggested that the Aah homologs are protein glycosyltransferases that modify their associated autotransporter, in a similar mechanism to that of Aah.

The Aah homologs, however, were not always associated with autotransporters that possess the 19 amino acid repeated consensus of AIDA-I. Indeed, in *E. coli* O127:H6 strain E2348/69 the monomeric autotransporter EhaJ, which does not possess the consensus motif, is glycosylated by its flanking *aah* homolog EgtA (Easton *et al.*, 2011). Furthermore, several *aah* homologs are flanked by putative trimeric autotransporters characterized by the conserved pfam03895 domain (Table 1). None of these trimeric autotransporters possess the 19 amino acid repeated consensus. In addition, three *aah* homologs (in *Acetobacter pasteurianus* IFO 3283-01, *Granulibacter bethesdensis* CGDNIH1 or *Bradyrhizobium sp.* ORS278) are flanked by putative adhesins or a large exoprotein, none of which seem related to the autotransporter family. And lastly, no autotransporter or adhesin genes were adjacent to eleven of the identified *aah* homologs. These observations suggest that Aah homologs may have diverse substrates and brings into question the substrate specificity of these enzymes. We therefore characterized the substrate specificity of Aah.

**Identification of a minimal peptide necessary and sufficient for glycosylation by Aah.**

In a previous study, we identified several residues in AIDA-I modified by heptose (Charbonneau et al., 2007). All these sites were located in the N-terminal region of the extracellular domain of AIDA-I comprising the imperfect repeats of 19 amino acids. Our strategy to determine the substrate specificity of Aah was to test whether small AIDA-I domains fused to GST could be glycosylated. A previously described fusion of amino acids A50 to G224 of AIDA-I (AIDAN) to GST encompasses at least 4 glycosylation sites and was glycosylated by Aah (Fig. 4A and D) (Charbonneau & Mourez, 2007). A smaller fragment of AIDA-I (amino acids A50 to G109, called AIDAN2), containing only one identified glycosylation site, was fused to GST (Fig. 4A). Whole cell lysates from bacteria co-expressing Aah and GST-AIDAN2 were prepared and probed with antibodies directed against GST to check expression of the fusion protein. Glycosylation was evaluated using antibodies against AIDA-I. As shown in Fig. 4D, GST-AIDAN2 was correctly expressed and glycosylated only in the presence of Aah.

We further engineered a GST-fusion containing only one repeat of 19 amino acids (amino acids N91 to G109, called Repeat). Each repeat is predicted to comprise three small  $\beta$ -strands, representing one turn of a right-handed  $\beta$ -helical structure (Fig. 4B and C) (Kajava & Steven, 2006). Again, by probing whole cell extracts with antibodies against GST or AIDA-I, we showed that the GST-Repeat fusion protein is correctly expressed and is glycosylated (Fig. 4D).

To confirm the glycosylation, we purified the GST-Repeat proteins in the presence or absence of Aah from the cytoplasm of the *E. coli* strain BL21 by affinity chromatography. Both proteins were digested with AspN and GluC endoproteases and samples were analysed by MS MALDI-TOF. As shown in Fig. 5, for the two samples, the DLVPRSGSNGGKTTATTVNSSGSQNVG peptide (with and without the N-terminal aspartate residue), corresponding to the Repeat sequence and few amino acids from the GST protein, were identified. However, for the glycosylated GST\_Repeat, additional peaks can be detected that are not present in the unglycosylated sample and that are consistent with a modification by one (increase of 192 Da) or two heptoses (increase of 384 Da). These results confirm that GST-Repeat is glycosylated and suggest that one or two heptose residues modify the 19 amino acid peptide.

To confirm that the glycosylation observed occurs on the AIDA-I peptide independently of the fusion partner, we also fused the same 19 amino acid sequence to a cytoplasmic variant



of the maltose binding protein (MalE<sub>C</sub>). As shown in Fig. 4E, the MalE<sub>C</sub>-Repeat fusion is correctly expressed at the expected size of 45 kDa. Furthermore, antibodies directed against AIDA-I revealed MalE<sub>C</sub>-Repeat, only in the presence of Aah. Taken together, our results showed that a 19 amino acid peptide is necessary and sufficient for glycosylation by Aah.

**Identification of the determinants required for glycosylation.** We replaced each amino acid of the Repeat peptide by an alanine to identify a motif required for recognition by Aah. The serine in position 12 (corresponding to S102 in native AIDA-I) is the amino acid we previously found to be modified (Charbonneau et al., 2007). Changing this serine into an alanine dramatically reduced glycosylation, although it did not prevent it completely (Fig. 6). This result confirms that Aah uses preferentially the serine 12 as an acceptor site. Glycosylation was altered to various extents in other mutants. Mutation of residues between positions N1-T9 or S15-G19 had no detectable effect. Modification of valine, asparagine, serine or glycine at position 10, 11, 13 and 14, respectively, caused a reduction in glycosylation efficiency but did not abolish it. Thus it seems unlikely that these amino acids are part of a strict recognition consensus. This would explain how Aah homologs could have substrates that do not possess the specific 19 amino acid consensus found in AIDA-I. Instead, Aah might recognize a specific conformation/structure.

Indeed, the Repeat peptide is predicted to adopt a secondary structure with three small  $\beta$ -strands, with the glycosylated serines in a loop between two strands (Fig. 4B). We speculated that Aah might recognize this conformation around the serine instead of a specific amino acid sequence. We therefore engineered new mutations that would modify the conformation of the Repeat peptide. All of the modified peptides were correctly expressed, although some were not glycosylated (Fig. 7). First, we deleted each of the predicted  $\beta$ -strands. The GR2-3 mutants, lacking the predicted  $\beta$ 2 or  $\beta$ 3 strands, were still glycosylated. These mutants maintained the predicted structural motif  $\beta$ -strand - acceptor loop -  $\beta$ -strand. However, the GR1 mutant lacking the predicted  $\beta$ 1 strand following the acceptor loop could not be glycosylated. It was not the sequence of the predicted  $\beta$ 1 strand that was important, since in the GR4 mutant we could swap this sequence with that of the predicted  $\beta$ 2 strand without affecting glycosylation. Thus, mutants GR1-4 suggest that a 'short  $\beta$ -strand - short acceptor loop - short  $\beta$ -strand' motif might be recognized by Aah.

We made a second series of mutations affecting the predicted loop. Mutant GR5 showed that shortening the predicted loop abolished glycosylation. It is possible that the loop in this

mutant might be destroyed altogether. Mutants GR6-7 demonstrated that adding an alanine at the beginning of the loop also had a dramatic effect on glycosylation, irrespective of the sequence of the preceding predicted strand. Mutant GR8 showed that adding an alanine at the end of the loop only slightly altered glycosylation. Thus, mutants GR5-8 suggest that Aah recognizes a short loop starting with the serine residue that is to be modified.

Most of mutants changed the 19 amino acid sequence length. In a last mutant, we aimed at recapitulating the putative structural requirements with a single mutation in the predicted  $\beta$ 3 strand. We hypothesized that the presence of a glycine, an amino acid rarely found in  $\beta$ -strand, would destroy the strand located before the serine residue. Indeed, when the valine residue of the predicted  $\beta$ 3 strand was changed by a glycine in mutant GR9, we observed that glycosylation was almost abolished.

Taken together, these observations suggest that the primary amino acid sequence is not the main signal recognized by Aah. Instead, Aah might recognize any sequence forming a 'short  $\beta$ -strand - short acceptor loop - short  $\beta$ -strand' motif, with the loop starting with a serine residue that will be targeted for glycosylation.

**Glycosylation of *Bordetella pertussis* pertactin by Aah.** The crystal structure of the *Bordetella pertussis* pertactin autotransporter shows that its extracellular domain folds into a right-handed  $\beta$ -helix (Emsley *et al.*, 1996). Thus, it comprises a succession of the 'short  $\beta$ -strand - short loop - short  $\beta$ -strand' motifs, several of which contain serine residues in the loop, but none of these motifs are formed by the 19 amino acid sequence consensus characteristic of AIDA-I, TibA or Ag43. Thus, we reasoned that if our hypothesis is correct, Aah should be able to glycosylate the extracellular domain of pertactin.

We constructed a signal sequence-less version of this domain (amino acids 35 to 563 of native pertactin), which we called pertactin\*. We added a six histidine tag at the N terminus of pertactin\* to allow detection. By probing whole cell lysates with antibodies directed against the six histidine tag or against glycosylated AIDA-I, we observed that pertactin\* is not only correctly expressed in these conditions but also glycosylated, only in the presence of Aah (Fig. 8). The ProQ Emerald 300 glycoprotein gel stain kit confirmed the glycosylation. These results are consistent with our mutagenesis study and support the idea that Aah does not recognize specific amino acids, but rather that Aah specificity relies on structural cues. Taken together, these results help explain the mechanism by which Aah homologs can glycosylate a diversity of substrates, with no apparent conserved primary

sequence consensus, by recognizing conserved structural aspects that are required for glycosylation among a diversity of bacterial species.

## Discussion

In this study, we have identified a new family of bacterial *O*-glycosyltransferases. Our BLAST search revealed that homologs of Aah can be found in various Gram-negative bacteria. We observed that high molecular weight proteins of *B. vietnamensis*, *C. sakazakii* and *C. rodentium* are indeed glycosylated. For several homologs, a predicted autotransporter protein with homologies to AIDA-I is located immediately adjacent to the glycosyltransferase gene, suggesting that it might be the substrate of the Aah homolog. We confirmed this assumption using chromosomal deletions and plasmid complementation in *C. rodentium*.

For all *Burkholderia* *ssp.*, except for *B. sp.* CCGE1003 and *B. vietnamensis*, we noted that Aah homologs are associated with a putative trimeric autotransporter. The YP\_105473 protein of *B. mallei*, identified in this study as an Aah homolog, has been previously named BimC and is found in a five gene operon encoding the trimeric autotransporter BimA (Schell *et al.*, 2007). BimA is required for the *Burkholderia* actin-based intracellular motility. The BimC protein was shown to be necessary for BimA-dependent actin tail formation. More interestingly, the authors have suggested that BimC may use sugar from the capsule biosynthetic pathway (6-deoxy- $\beta$ -D-manno-heptopyranose) for potential glycosylation of BimA, which is reminiscent of the use of the LPS biosynthetic pathway by Aah. These observations reinforce our hypothesis that Aah homologs also act as heptosyltransferases. Glycosylation of a trimeric autotransporter, however, is unusual. To the best of our knowledge, only the EmaA collagen adhesin of *Aggregatibacter actinomycetemcomitans*, a human oropharyngeal pathogen, has been described as a glycosylated trimeric autotransporter (Tang & Mintz, 2010). Interestingly, for this protein, sugars used for glycosylation also originate from the LPS biosynthetic pathway. Our findings suggest that this new family of *O*-glycosyltransferases can also be associated with glycosylation of trimeric autotransporters and it would be interesting to test this hypothesis in the future.

Strikingly, the trimeric autotransporters associated with an Aah homolog do not possess the 19 amino acid consensus that seemed to be specifically recognized by Aah or TibC. Other monomeric autotransporters or adhesins associated with an Aah homolog did

not have the consensus either. This observation suggested that this consensus may not be required for efficient glycosylation by Aah. By undertaking a study to determine the substrate specificity of Aah, we came to the conclusion that Aah might recognize a small structural motif of 'short  $\beta$ -strand - short loop - short  $\beta$ -strand', with the short acceptor loop starting with a serine that is modified. This structural motif is often comprised in right-handed  $\beta$ -helices characteristic of monomeric autotransporters or in the two-stranded  $\beta$ -rolls found in trimeric autotransporters, which could explain why Aah homologs are most often associated with autotransporters (Kajava & Steven, 2006).

For almost all studied *O*-glycosyltransferases, identification of a precise primary amino acid consensus has failed, like for Aah. Some authors have already suggested that *O*-glycosyltransferases may recognize an element of secondary structure. This is the case of the *N*-acetylgalactosaminyltransferases involved in the mucin-type *O*-linked glycosylation in eukaryotes (Gerken *et al.*, 1997). For this protein, an extended  $\beta$ -like conformation seems favorable for the enzyme-substrate binding. The recent crystal structure of a human *O*-linked  $\beta$ -*N*-acetylglucosamine transferase (OGT) in complex with a peptide substrate also suggests that this glycosyltransferase recognized its cognate substrate in an extended conformation, with a preference for sequences in which the residues flanking the glycosylated amino acid enforce an extended conformation, like proline and  $\beta$ -branched amino acids (Lazarus *et al.*, 2011).

For AIDA-I, glycosylation has been shown to be essential to ensure the normal conformation and the adhesion mediated by this protein (Benz & Schmidt, 2001, Charbonneau *et al.*, 2007). Similarly, glycosylation of Ag43 protects the protein against thermal and chemical denaturation, and addition of sugars is required for optimal binding of the trimeric autotransporter EmaA to collagen (Knudsen *et al.*, 2008, Tang & Mintz, 2010). Therefore, these studies demonstrate that glycosylation plays important structural and/or functional roles for bacterial autotransporters proteins. The finding that other autotransporters or large adhesin encoding genes are linked to Aah homolog encoding genes suggests that glycosylation may be important for the function of these proteins, as for AIDA-I, Ag43 and EmaA.

Another family of bacterial oligosaccharyltransferases, represented by the HMW1C protein of *Haemophilus influenzae*, has been described recently (Grass *et al.*, 2010). This alternative bacterial pathway for glycosylation presents a mechanism of action similar to the one of the Aah family, since both pathways involve a unique cytoplasmic

glycosyltransferase that uses nucleotide-activated monosaccharides as sugar donors (Grass *et al.*, 2003, Gross *et al.*, 2008). Furthermore, the HMW1c family also modifies bacterial surface protein secreted via the type V secretion pathway and glycosylation of the HMW1A protein of *H. influenzae* is required to prevent premature degradation and efficient tethering of the protein at the cell surface, like AIDA-I (Grass *et al.*, 2003). However, there are major differences between these two families. The HMW1c family proteins are *N*-glycosyltransferases that recognize the conventional consensus sequon N\*- X- S/T (Gross *et al.*, 2008). Moreover, HMW1c-like proteins can also have an *O*-glycosyltransferase function as they created di-hexose complexes on the acceptor protein, an activity never described for Aah.

Glycosylation of proteins is now recognized as an important modification in bacteria and opens up the possibility of protein glycoengineering in prokaryotes. However, even as more information becomes available about bacterial glycosylation, unanswered questions remain. In this study, we bring new information about the mechanism of action of a unique cytoplasmic *O*-glycosyltransferase of *E. coli*. We found that Aah can recognize a peptide as short as 13 amino acids, when fused to a heterologous protein. This suggests that it could be possible to recombinantly target other proteins from prokaryotes or even eukaryotes for *O*-glycosylation by adding this short peptide sequence, without any other requirements. Of note, this modification is compatible with the bacterial secretion machinery. Thus, this result opens new possibilities for glycoengineering applications in *E. coli*.

## **Experimental procedures**

### **Bacterial strains**

The *Escherichia coli* laboratory strains C600 (F<sup>-</sup> *thr-1 leuB6 thi-1 lacY1 supE44 rfbD1 fhuA21*), BL21 (F<sup>-</sup> *ompT hsdS*(r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>) *gal*) and ED9 (*araD ΔlacU169 rpsL relA flbB deoC ptsF ΔmalE444*) were used in this study. We also used the previously described pathogenic strain *E. coli* 2787 (Benz & Schmidt, 1989). *Citrobacter rodentium* DBS100, *Burkholderia vietnamiensis* G4, *Cronobacter sakazakii* HBP 2871 and HPB 3196 were kindly provided by Samantha Gruenheid, McGill University, Éric Déziel, INRS-Institut Armand-Frappier and Franco Pagotto, Health Canada, respectively.

### **Construction of plasmids and bacterial recombinant strains**

Plasmids pAah and pGEX-AIDAN enabling the expression, respectively, of Aah and GST-AIDAN have been described previously (Charbonneau & Mourez, 2007). For construction of the pGEX\_AIDAN2 and pGEX-Repeat plasmids and for introducing point mutations or deletions in the pGEX-Repeat plasmid, we used the QuikChange II site-directed mutagenesis kit (Stratagene) using the primers listed in Table S1 and their corresponding complementary oligonucleotides. For amplification of the 19 amino acid peptide, we used primers MalE\_Repeat\_F and MalE\_Repeat\_R (Table S1), which introduced recognition sites for restriction enzymes *EcoRI* and *XbaI*. The resulting fragment was cloned in the pMalp2x vector (New England Biolabs), generating pMalE-Repeat. This plasmid was treated with the *BglII* and *XbaI* restriction enzymes and the resulting fragment was introduced in the pMal-c4X vector (New England Biolabs). The resulting pMalc-Repeat plasmid allows expression of the MalE-Repeat fusion protein in the cytoplasm of the bacteria. The *aidA*-like gene from *C. sakazakii* (ESA\_02084) was cloned from *C. sakazakii* HPB 3196. The gene was amplified by PCR in two fragments using the primers AIDA<sub>Cs</sub>\_F/AIDA<sub>Cs</sub>\_R and AIDA<sub>Cs</sub>\_F2/AIDA<sub>Cs</sub>\_R2 listed in Table S1. Amplification products were digested by *NcoI/SmaI* and *SmaI/XbaI*, respectively and introduced in the pTRC99a vector (Pharmacia Biotech). The resulting plasmid was named pAIDA<sub>Cs</sub>. For construction of the pAah<sub>Cr</sub> plasmid, the *aah*-like gene was amplified from a *C. rodentium* DBS100 lysate using primers Aah<sub>Cr</sub>\_F and Aah<sub>Cr</sub>\_R (Table S1). The resulting PCR fragment was digested with *PciI* and *XhoI* and cloned into the *NcoI* and *SalI* restriction sites of the pTRC99a vector. The pPert $\Delta$ SS $\Delta$  $\alpha\beta$  plasmid was obtained by amplification of the region corresponding to amino acids 35 to 563 of native Pertactin of *Bordetella pertussis* from the pRNAC plasmid (kindly provided by Ian R. Henderson, University of Birmingham) with primers Pertactin\_F and Pertactin\_R (Table S1). The PCR fragment was digested with the restriction enzymes *NcoI* and *XbaI* and cloned in the pTRC99a vector. For introduction of the N-terminal six histidine tag, we used the QuikChange II site-directed mutagenesis kit with primer pPertHis and its corresponding complementary oligonucleotide, thus generating pPertHis $\Delta$ SS $\Delta$  $\alpha\beta$  (expressing Pertactin\* protein). In all constructs, protein expression is under the control of isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) through the *ptac* promoter. All constructions were verified by restriction mapping and sequencing.

The isogenic mutant 2787  $\Delta$ *aidA* was constructed by allelic exchange using the pAIDA::Gm plasmid carrying the  $\Delta$ *aidA*::*aacCI* allele (Ravi *et al.*, 2007). The plasmid was transformed in *E. coli* strain SM10 (*thr leu tonA lacy supE recA*::RP4-2-Tc::Mu Km<sup>R</sup>  $\lambda$  *pir*)

by electroporation, then transferred to *E. coli* 2787 recipient cell by conjugation. After incubation at 37°C overnight on blood agar plates, the exconjugants were selected on M9 agar plate (without thiamine) containing ampicillin (100 µg ml<sup>-1</sup>). A few colonies were spread onto LB agar containing 10% sucrose and gentamycin (20 µg ml<sup>-1</sup>) and incubated at 30°C overnight in order to promote a double crossover. A few sucrose- and Gm-resistant colonies were selected and the mutation was confirmed by PCR using primers Exc\_AIDA-F and exc\_AIDA-R.

*Citrobacter rodentium* DBS100 mutants were constructed by using the procedure described by Datsenko and Wanner (Datsenko & Wanner, 2000). The pKD13 vector was used for the kanamycin resistance cassette. Antibiotic insertions were removed by transforming the mutant strains with the pCP20 plasmid expressing the FLP recombinase. Primers used for generation of these mutants are listed in Table S1.

### **Bacterial growth and induction conditions**

Bacteria containing the different plasmids were grown on Luria-Bertani (LB) agar plates or in liquid LB medium containing 100 µg ml<sup>-1</sup> of ampicillin, and, in addition, 50 µg ml<sup>-1</sup> chloramphenicol when the plasmid pAah was used. Bacterial cultures were grown at 30°C until an optical density at 600 nm (O.D.<sub>600nm</sub>) of 0.4 was reached and then were induced with 10 µM of IPTG for 3 hours. For expression of pAIDA<sub>Cs</sub> plasmid, bacteria were grown at 30°C and induced overnight with 10 µM of IPTG following growth to an O.D.<sub>600nm</sub> of 0.8. The wild type strains *E. coli* 2787, *C. rodentium* DBS100, *B. vietnamiensis* G4 and *C. sakazakii* HBP 2871, were grown overnight at 37°C.

### **Whole cell extracts**

Bacterial cultures were grown and induced as described previously, normalized to the same O.D.<sub>600nm</sub> and centrifuged for 15 minutes at 12,000 x g in microfuge tubes. The pellets were resuspended in 50 µl of Phosphate Buffered Saline (PBS). All samples were processed by SDS-PAGE and immunoblotting, as described below.

### **SDS-PAGE, immunoblotting and glycan detection**

Protein-containing samples were diluted in twice concentrated SDS-PAGE loading buffer containing β-mercaptoethanol and denatured by heating at 100°C for 10 minutes. The samples were then separated by SDS-PAGE on 7% or 12% acrylamide gels. The gels were

either stained with Coomassie blue or transferred to polyvinylidene fluoride membranes (PVDF, Millipore). Alternatively, for immunodot blot, denatured proteins were directly immobilized on a PVDF membrane. Immunodetection was performed with a serum raised against a glycosylated heat-extracted portion of the extracellular domain of AIDA-I, diluted 1:10,000 or 1:5,000 in blocking buffer (5% skim milk, 50 mM Tris-HCl pH 8, 150 mM NaCl, 0.05% Triton X100). A goat anti-rabbit horseradish peroxidase (HRP)-coupled antibody (Sigma) diluted 1:10,000 was used as the secondary antibody, according to the manufacturer's instructions. Alternatively, anti-GST antibodies diluted 1:10,000 were used for detection of the GST-fusion proteins. For detection of the MalE-Repeat fusion protein, Anti-MalE antibodies (New England Biolabs) diluted 1:10,000 was used. Immune complexes were revealed using a 3,3',5,5'-tetramethylbenzidine solution for membranes (Sigma). The detection of glycosylation was accomplished using the ProQ Emerald 300 glycoprotein gel and blot stain kit (Molecular Probes), according to the instructions of the manufacturer.

### **Purification of the GST-Repeat protein**

One liter of *E. coli* BL21 harboring plasmids pGex-Repeat with or without pAah were grown until an O.D.<sub>600nm</sub> of 0.4 and induced with 10  $\mu$ M of IPTG for three hours. Bacteria were harvested and resuspended in 50 ml of Tris-buffered saline (50 mM Tris-HCl pH 8, 150 mM NaCl) containing lysozyme (0.4 mg ml<sup>-1</sup> final) and lysed with French press and an ultrasonic processor. The soluble fraction was isolated by a 40 minute centrifugation at 16,000 x g. The GST and the GST-Repeat fusion proteins were purified using an ÄKTA purifier system with a 1-ml glutathione sepharose column (Amersham biosciences), according to instructions of the manufacturer. The purity of the proteins was confirmed by SDS-PAGE and staining with Coomassie blue. For the mass spectrometry experiment, 1  $\mu$ g of GST-repeat unglycosylated and glycosylated were digested overnight with AspN and GluC endoproteinases (New England Biolabs) in solution. The samples were dried, and peptide fragments were desalted using zip-tip<sub>C18</sub> (Millipore) according to the supplier protocol and dissolved in 0.1% formic acid. Samples were spotted on an MTP AnchorChip<sup>TM</sup> 600/384 target plate using CHCA (alpha-Cyano-4-hydroxycinnamic acid) as a matrix. The MALDI-TOF MS mass spectra were obtained on an UltrafleXtreme<sup>TM</sup> MALDI TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). LC-MS/MS



analysis was performed on a Waters Premier QTOF mass spectrometer interfaced with a Waters nanoAcquity UPLC System (Waters, Milford, MA, USA).

### Sequence alignment and phylogenetic tree construction

Protein sequences were obtained from the NCBI database (<http://www.ncbi.nlm.nih.gov/protein>) and blast search was performed using BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>). Sequence alignment was performed using the BLOSUM 62 matrix, with the Geneious Pro 4.0.2 software (Biomatters Ltd.). Only proteins showing an E-value  $\leq 10^{-73}$  were retained. Based on the alignment, a neighbour-joining tree with 1000 bootstrap replicates was computed using the Geneious Pro 4.0.2 software (Biomatters Ltd.).

### Acknowledgments

We thank Ian R. Henderson, Samantha Gruenheid, Éric Déziel and Franco Pagotto, for their generous gifts of strains and plasmids. This work was supported by financial contributions from the Groupe de Recherche et d'Études sur les Maladies Infectieuses du Porc (GREMIP), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair program to M. M and C. M. D. M.-È. C. received a NSERC graduate fellowship (Grant 85297596).

### References

- Benz, I. & M. A. Schmidt, (1989) Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infection and immunity* **57**: 1506-1511.
- Benz, I. & M. A. Schmidt, (2001) Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Molecular microbiology* **40**: 1403-1413.
- Chapman, T. A., X. Y. Wu, I. Barchia, K. A. Bettelheim, S. Driesen, D. Trott, M. Wilson & J. J. Chin, (2006) Comparison of virulence gene profiles of *Escherichia coli* strains isolated from healthy and diarrheic swine. *Applied and environmental microbiology* **72**: 4782-4795.

- Charbonneau, M. E., F. Berthiaume & M. Mourez, (2006) Proteolytic processing is not essential for multiple functions of the *Escherichia coli* autotransporter adhesin involved in diffuse adherence (AIDA-I). *Journal of bacteriology* **188**: 8504-8512.
- Charbonneau, M. E., V. Girard, A. Nikolakakis, M. Campos, F. Berthiaume, F. Dumas, F. Lepine & M. Mourez, (2007) O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence. *Journal of bacteriology* **189**: 8880-8889.
- Charbonneau, M. E. & M. Mourez, (2007) Functional organization of the autotransporter adhesin involved in diffuse adherence. *Journal of bacteriology* **189**: 9020-9029.
- Datsenko, K. A. & B. L. Wanner, (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 6640-6645.
- Dobos, K. M., K. H. Khoo, K. M. Swiderek, P. J. Brennan & J. T. Belisle, (1996) Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *Journal of bacteriology* **178**: 2498-2506.
- Easton, D. M., M. Totsika, L. P. Allsopp, M. D. Phan, A. Idris, D. J. Wurpel, O. Sherlock, B. Zhang, C. Venturini, S. A. Beatson, T. J. Mahony, R. N. Cobbold & M. A. Schembri, (2011) Characterization of EhaJ, a New Autotransporter Protein from Enterohemorrhagic and Enteropathogenic *Escherichia coli*. *Front Microbiol* **2**: 120.
- Emsley, P., I. G. Charles, N. F. Fairweather & N. W. Isaacs, (1996) Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature* **381**: 90-92.
- Fletcher, C. M., M. J. Coyne & L. E. Comstock, (2011) Theoretical and experimental characterization of the scope of protein O-glycosylation in *Bacteroides fragilis*. *The Journal of biological chemistry* **286**: 3219-3226.
- Fletcher, C. M., M. J. Coyne, O. F. Villa, M. Chatzidaki-Livanis & L. E. Comstock, (2009) A general O-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell* **137**: 321-331.
- Gerken, T. A., C. L. Owens & M. Pasumarthy, (1997) Determination of the site-specific O-glycosylation pattern of the porcine submaxillary mucin tandem repeat glycopeptide. Model proposed for the polypeptide:galnac transferase peptide binding site. *The Journal of biological chemistry* **272**: 9709-9719.
- Gerwig, G. J., J. P. Kamerling, J. F. Vliegthart, E. Morag, R. Lamed & E. A. Bayer, (1993) The nature of the carbohydrate-peptide linkage region in glycoproteins from

- the cellulosomes of *Clostridium thermocellum* and *Bacteroides cellulosolvens*. *The Journal of biological chemistry* **268**: 26956-26960.
- Grass, S., A. Z. Buscher, W. E. Swords, M. A. Apicella, S. J. Barenkamp, N. Ozchlewski & J. W. St Geme, 3rd, (2003) The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Molecular microbiology* **48**: 737-751.
- Grass, S., C. F. Lichti, R. R. Townsend, J. Gross & J. W. St Geme, 3rd, (2010) The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathog* **6**: e1000919.
- Gross, J., S. Grass, A. E. Davis, P. Gilmore-Erdmann, R. R. Townsend & J. W. St Geme, 3rd, (2008) The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. *The Journal of biological chemistry* **283**: 26010-26015.
- Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. C. Fernandez & D. Ala'Aldeen, (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* **68**: 692-744.
- Ielmini, M. V. & M. F. Feldman, (2011) *Desulfovibrio desulfuricans* PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* **21**: 734-742.
- Kajava, A. V. & A. C. Steven, (2006) The turn of the screw: variations of the abundant beta-solenoid motif in passenger domains of Type V secretory proteins. *Journal of structural biology* **155**: 306-315.
- Klemm, P., R. M. Vejborg & O. Sherlock, (2006) Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol* **296**: 187-195.
- Knudsen, S. K., A. Stensballe, M. Franzmann, U. B. Westergaard & D. E. Otzen, (2008) Effect of glycosylation on the extracellular domain of the Ag43 bacterial autotransporter: enhanced stability and reduced cellular aggregation. *The Biochemical journal* **412**: 563-577.
- Kowarik, M., N. M. Young, S. Numao, B. L. Schulz, I. Hug, N. Callewaert, D. C. Mills, D. C. Watson, M. Hernandez, J. F. Kelly, M. Wacker & M. Aebi, (2006) Definition of

- the bacterial N-glycosylation site consensus sequence. *The EMBO journal* **25**: 1957-1966.
- Lazarus, M. B., Y. Nam, J. Jiang, P. Sliz & S. Walker, (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**: 564-567.
- Lindenthal, C. & E. A. Elsinghorst, (1999) Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infection and immunity* **67**: 4084-4091.
- Moormann, C., I. Benz & M. A. Schmidt, (2002) Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infection and immunity* **70**: 2264-2270.
- Niewerth, U., A. Frey, T. Voss, C. Le Bouguenec, G. Baljer, S. Franke & M. A. Schmidt, (2001) The AIDA autotransporter system is associated with F18 and stx2e in *Escherichia coli* isolates from pigs diagnosed with edema disease and postweaning diarrhea. *Clinical and diagnostic laboratory immunology* **8**: 143-149.
- Nothaft, H. & C. M. Szymanski, (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**: 765-778.
- Plummer, T. H., Jr., A. L. Tarentino & C. R. Hauer, (1995) Novel, specific O-glycosylation of secreted *Flavobacterium meningosepticum* proteins. Asp-Ser and Asp-Thr-Thr consensus sites. *The Journal of biological chemistry* **270**: 13192-13196.
- Ravi, M., M. Ngeleka, S. H. Kim, C. Gyles, F. Berthiaume, M. Mourez, D. Middleton & E. Simko, (2007) Contribution of AIDA-I to the pathogenicity of a porcine diarrheagenic *Escherichia coli* and to intestinal colonization through biofilm formation in pigs. *Veterinary microbiology* **120**: 308-319.
- Schell, M. A., R. L. Ulrich, W. J. Ribot, E. E. Brueggemann, H. B. Hines, D. Chen, L. Lipscomb, H. S. Kim, J. Mrazek, W. C. Nierman & D. Deshazer, (2007) Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Molecular microbiology* **64**: 1466-1485.
- Schmidt, M. A., L. W. Riley & I. Benz, (2003) Sweet new world: glycoproteins in bacterial pathogens. *Trends in microbiology* **11**: 554-561.
- Sherlock, O., U. Dobrindt, J. B. Jensen, R. Munk Vejborg & P. Klemm, (2006) Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. *Journal of bacteriology* **188**: 1798-1807.

- Sherlock, O., M. A. Schembri, A. Reisner & P. Klemm, (2004) Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm formation. *Journal of bacteriology* **186**: 8058-8065.
- Szymanski, C. M. & B. W. Wren, (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* **3**: 225-237.
- Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust & P. Guerry, (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular microbiology* **32**: 1022-1030.
- Tang, G. & K. P. Mintz, (2010) Glycosylation of the collagen adhesin EmaA of *Aggregatibacter actinomycetemcomitans* is dependent upon the lipopolysaccharide biosynthetic pathway. *Journal of bacteriology* **192**: 1395-1404.
- Vik, A., F. E. Aas, J. H. Anonsen, S. Bilsborough, A. Schneider, W. Egge-Jacobsen & M. Koomey, (2009) Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 4447-4452.
- Yan, Q. & W. J. Lennarz, (1999) Oligosaccharyltransferase: a complex multisubunit enzyme of the endoplasmic reticulum. *Biochem Biophys Res Commun* **266**: 684-689.
- Zhang, W., M. Zhao, L. Ruesch, A. Omot & D. Francis, (2007) Prevalence of virulence genes in *Escherichia coli* strains recently isolated from young pigs with diarrhea in the US. *Veterinary microbiology* **123**: 145-152.

## Figure legends

### Fig. 1: Phylogenetic tree of Aah homologs.

Neighbor-joining tree based on 19 Aah homologs. The *E. coli* WaaC and RfaF heptosyltransferases involved in LPS biosynthesis were added as an outgroup (sequence identities with Aah of 12.9 % for WaaC and 15.1 % for RfaF). Lines are colored based on the taxonomy of the related organisms: red for  $\gamma$ -proteobacteria, blue for  $\alpha$ -proteobacteria and green for  $\beta$ -proteobacteria. The name, the organism and the predicted function of the associated open reading frame are indicated (M: monomeric autotransporter, T: trimeric autotransporter, U: other proteins unrelated to the autotransporter family, N/D: no data) for each Aah homologs. The lengths of the branches reflect the number of substitutions per site

between different sequences. Bootstrap values from 1000 replicates are shown for each branch points.

**Fig. 2: Glycosylation in various bacterial species**

*A.* Immunodot blot with antibody against glycosylated AIDA-I was performed on whole cell lysates obtained from overnight cultures of *E. coli* 2787, *E. coli* 2787  $\Delta aidA$ , *C. rodentium* DBS100, *C. sakazakii* HBP 2871 and *B. vietnamensis* G4. *B.* Whole cell lysates were obtained from overnight cultures of *E. coli* 2787, *E. coli* 2787  $\Delta aidA$ , *C. rodentium*, *C. rodentium*  $\Delta aah_{Cr}$ , *C. rodentium*  $\Delta aidA_{Cr}$  and *C. rodentium*  $\Delta aah_{Cr}$  bearing the pAah<sub>Cr</sub> plasmid. Proteins separated by SDS-PAGE were probed by immunoblotting with antibodies directed against glycosylated AIDA-I.

**Fig. 3: Glycosylation of a *Cronobacter sakazakii* protein by Aah.**

*A.* Schematic representation of the AIDA-I protein and the putative *C. sakazakii* autotransporter protein (AIDA<sub>CS</sub>), showing the N-terminal signal sequence, a region encompassing the imperfect repeat of 19 amino acids (35 repetitions for AIDA-I and 93 for AIDA<sub>CS</sub>, represented by the light grey boxes) and the C-terminal membrane-embedded domain. The Weblogo (<http://weblogo.berkeley.edu/>) representation of the repeated motif found in AIDA-I and AIDA<sub>CS</sub> show the relative conservation of each amino acid of the consensus. *B.* Whole cell lysates were obtained from cultures of *E. coli* strain C600 bearing the control vector (pTRC99a, lane 1) or the pAIDA<sub>CS</sub> plasmid (lanes 2 and 3). The presence of the pAah plasmid is highlighted (lane 3). Proteins separated by SDS-PAGE were stained with Coomassie blue or probed with antibodies directed against glycosylated AIDA-I. Detection of glycans was performed using the ProQ Emerald glycoprotein gel and blot stain kit.

**Fig. 4: Identification of a short sequence sufficient for glycosylation by Aah.**

*A.* Schematic representation of the different fusions of the N-terminal domain of AIDA-I to GST. *B.* Representation of the coil model for AIDA-I. A filled circle represents an isoleucine, valine, or leucine residue, and “X” indicates any residue. *C.* Schematic representation of the 19 amino acid peptide fused to the GST protein. An arrowhead indicates the position of the previously identified glycosylated serine. Each repeat of this motif is thought to correspond to a coil of a  $\beta$ -helix and is predicted to consist of three

small  $\beta$ -strands. *D.* Whole cell lysates were obtained from cultures of *E. coli* strain C600 bearing control vector (pGex-4-T1), pGex-Repeat, pGex-AIDAN or the pGex-AIDAN2 plasmid. The presence of the pAah plasmid is highlighted. Proteins separated by SDS-PAGE were probed by immunoblotting with antibodies directed against the GST protein or against glycosylated AIDA-I. *E.* Schematic representation of the 19 amino acid peptide (Repeat) fused to the MalE<sub>C</sub> protein. Whole cell extracts of *E. coli* strain ED9 harboring the control vector (pMal-c4X) and the pAah plasmid, the pMalEc\_Repeat plasmid or the pMalEc\_Repeat and pAah plasmids were separated by SDS-PAGE and were probed by immunoblotting with antibodies directed against the MalE protein or against glycosylated AIDA-I.

**Fig. 5: Glycosylation of the GST-Repeat protein.**

Purified glycosylated (top panel) and unglycosylated (bottom panel) GST-Repeat proteins were digested by AspN and GluC endoproteinases and analyzed by MS MALDI-TOF. Peaks corresponding to the Repeat peptide are identified. Hep: heptose, D: aspartate residue.

**Fig. 6: Alanine-scanning of the 19 amino acid sequence.**

Whole cell lysates were obtained from cultures of *E. coli* strain BL21 bearing pAah and the pGex-Repeat plasmid (WT) or one of the plasmids bearing a single mutation in the pGex-Repeat. Proteins separated by SDS-PAGE were probed by immunoblotting with antibodies directed against the GST protein or against glycosylated AIDA-I (upper panel). The intensity of the bands probed with the AIDA-I antibody was quantitated using the ImageJ software and normalized (lower panel). The values obtained from three different experiments represent means  $\pm$  standard errors of the means. Significant differences are indicated by asterisks (\*,  $P < 0.1$ , \*\*,  $P < 0.05$ , \*\*\*  $P < 0.001$ ).

**Fig. 7: Deletion mutagenesis of the GST-Repeat protein.**

Schematic representations of the different GST-Repeat mutants (upper panel, left). The predicted glycosylated serine is underlined. An arrow represents the predicted  $\beta$ -strands for each construction. Whole cell lysates were obtained from cultures of *E. coli* strain BL21 bearing pAah and the pGex-Repeat plasmid (WT) or one of the plasmids bearing a mutation in pGex-Repeat. Proteins separated by SDS-PAGE were probed by

immunoblotting with antibodies directed against the GST protein or against glycosylated AIDA-I (upper panel, right). The intensity of the bands probed with the AIDA-I antibody was quantitated using the ImageJ software and normalized (lower panel). The values obtained from three different experiments represent means  $\pm$  standard errors of the means. Significant differences were indicated by asterisks (\*,  $P < 0.1$ , \*\*,  $P < 0.05$ , \*\*\*  $P < 0.001$ ).

**Fig. 8: Glycosylation of the *B. pertussis* pertactin autotransporter protein.**

Whole cell lysates were obtained from cultures of *E. coli* strain BL21 bearing the control vector (pTRC99a) or the pPertactin\* plasmid in the presence or absence of the pAah plasmid, as indicated. Proteins separated by SDS-PAGE were probed with antibodies directed against the Pertactin\*His-tag or against glycosylated AIDA-I. Detection of glycans from whole cell lysates was performed using the ProQ Emerald glycoprotein stain kit.



Table I: Predicted glycosyltransferase encoding open reading frames (ORFs) with homology to Aah identified in bacterial genomes

Accession no.	Length		Organism	Associated ORF	
	(a.a)	% id		Accession no.	Putative function
<i>Gammaproteobacteria</i>					
CAAC43407	391	100 (391)	<i>Escherichia coli</i> str. 2787	CAA46156	AIDA-1 (*)
ABS20375	391	99 (391)	<i>Escherichia coli</i>	N/D	N/D
ZP_06661292	409	69 (385)	<i>Escherichia coli</i> BO88	ZP_06661293	monomeric AT (*)
Q9S4K6	406	68 (385)	<i>Escherichia coli</i> H10407	AAD41751	TibA (*)
ZP_07522324	406	67 (388)	<i>Escherichia coli</i> TA271	ZP_07522323 <sup>§</sup>	monomeric AT
YP_003033978	406	67 (388)	<i>Escherichia coli</i> Vir68	YP_003033977	monomeric AT (*)
YP_002330203	411	62 (391)	<i>Escherichia coli</i> O127:H6 str. E2348/69	YP_002330202	EhaJ
YP_003368456	409	69 (385)	<i>Citrobacter rodentium</i> ICC168	YP_003368457	monomeric AT (*)
ZP_05433042	406	67 (388)	<i>Shigella</i> sp. D9	-	-
ZP_05727241	410	67 (385)	<i>Pantoea</i> sp. At-9b	ZP_05727242	monomeric AT (*)
YP_001438171	420	60 (390)	<i>Cronobacter sakazakii</i> ATCC BAA-894	YP_001438170	monomeric AT (*)
YP_003210243	430	60 (390)	<i>Cronobacter turicensis</i> z3032	YP_003210244	monomeric AT (*)
<i>Alphaproteobacteria</i>					
YP_001208059	435	53 (387)	<i>Bradyrhizobium</i> sp. ORS278	YP_001208060	large exoprotein
YP_001205949	709	48 (382)	<i>Bradyrhizobium</i> sp. ORS278	-	-
YP_001240290	711	48 (382)	<i>Bradyrhizobium</i> sp. BTA11	-	-
YP_744277	465	51 (398)	<i>Granulibacter thebesdensis</i> CGDNIH1	-	-
YP_745467	427	47 (394)	<i>Granulibacter thebesdensis</i> CGDNIH1	YP_745466	adhesin family
YP_003186962	428	51 (376)	<i>Acetobacter pasteurianus</i> IFO 3283-01	-	-
YP_003186686	423	48 (391)	<i>Acetobacter pasteurianus</i> IFO 3283-01	YP_003186685	adhesin family
YP_191325	415	46 (398)	<i>Gluconobacter oxydans</i> 621H	-	-
YP_191427	349	45 (351)	<i>Gluconobacter oxydans</i> 621H	-	-
<i>Betaproteobacteria</i>					
ZP_04578853	418	62 (381)	<i>Oxalobacter formigenes</i> OXCC13	N/D	N/D
ZP_04575883	461	61 (377)	<i>Oxalobacter formigenes</i> HOxBLS	N/D	N/D
YP_002795895	398	50 (374)	<i>Laribacter hongkongensis</i> HLHK9	YP_002795894	monomeric AT (*)

ZP_06468799	444	53 (384)	<i>Burkholderia</i> sp. CCGE1003	ZP_06468800	monomeric AT (*)
YP_001110605	452	48 (397)	<i>Burkholderia vietnamiensis</i> G4	YP_001110606	monomeric AT (*)
ZP_02384278	441	42 (382)	<i>Burkholderia thailandensis</i> Bt4	-	-
YP_439073	423	42 (382)	<i>Burkholderia thailandensis</i> E264	YP_439072	trimeric AT
ZP_02370369	423	42 (382)	<i>Burkholderia thailandensis</i> TXDOH	-	-
ZP_02466797	409	41 (381)	<i>Burkholderia thailandensis</i> MSMB43	N/D	N/D
ZP_00438906	453	41 (383)	<i>Burkholderia mallei</i> GB8 horse 4	ZP_00438905	trimeric AT
ZP_04915511	470	41 (383)	<i>Burkholderia mallei</i> JHU	ZP_04915510	trimeric AT
YP_105473	406	41 (383)	<i>Burkholderia mallei</i> ATCC 23344	YP_105472	trimeric AT
ZP_02269538	367	42 (366)	<i>Burkholderia mallei</i> PRL-20	ZP_04820410	trimeric AT
YP_001063111	453	41 (383)	<i>Burkholderia pseudomallei</i> 668	ZP_03456696	trimeric AT
ZP_01768658	466	42 (376)	<i>Burkholderia pseudomallei</i> 305	ZP_01768706	trimeric AT
ZP_04522854	466	42 (376)	<i>Burkholderia pseudomallei</i> MSHR346	N/D	N/D
YP_001076052	466	42 (376)	<i>Burkholderia pseudomallei</i> 1106a	YP_001076053	trimeric AT
YP_335685	466	42 (376)	<i>Burkholderia pseudomallei</i> 1710b	YP_335687	trimeric AT
ZP_03791705	466	42 (376)	<i>Burkholderia pseudomallei</i> Pakistan 9	ZP_03791711	trimeric AT
ZP_04893325	466	42 (376)	<i>Burkholderia pseudomallei</i> Pasteur 52237	ZP_04893326	trimeric AT
YP_111498	419	42 (376)	<i>Burkholderia pseudomallei</i> K96243	YP_111499	trimeric AT
ZP_03456680	466	42 (376)	<i>Burkholderia pseudomallei</i> 576	ZP_03456696	trimeric AT
ZP_02360032	439	40 (386)	<i>Burkholderia oklahomensis</i> EO147	-	-
ZP_02366136	439	41 (380)	<i>Burkholderia oklahomensis</i> C6786	-	-

The accession number, the length of the ORF in amino acids (a.a), the percentage of identity and the number of amino acids corresponding (% id (a.a)) and the organism are indicated for each Aah homolog. The presence of a predicted autotransporter (AT) or adhesin protein next to the putative glycosyltransferase has been noted by the accession number, the putative function and the presence of the specific 19 amino acid consensus (\*). § indicates that the ORF is prematurely interrupted whereas N/D refers to the absence of information about the region surrounding the glycosyltransferase encoding

Figure 1

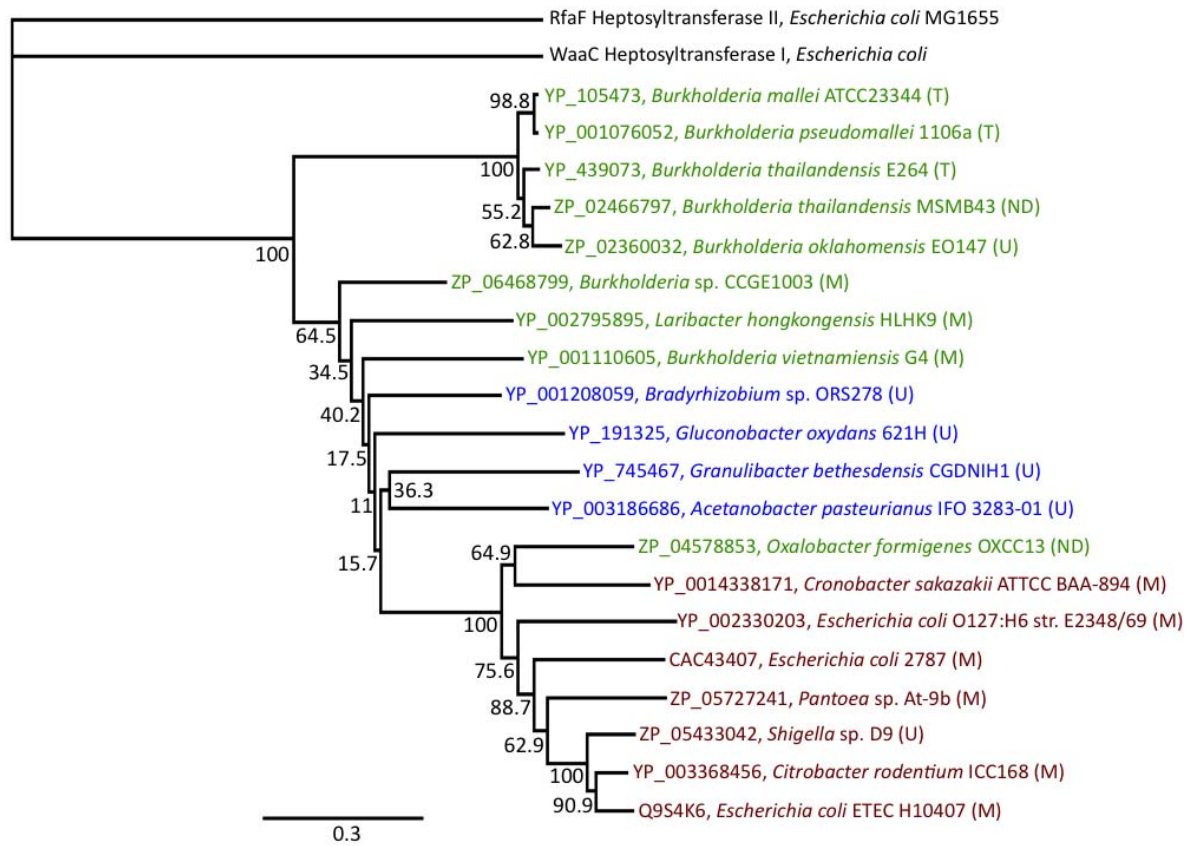


Figure 2

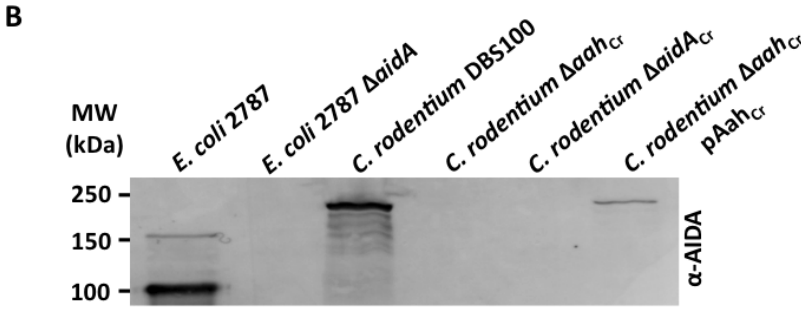
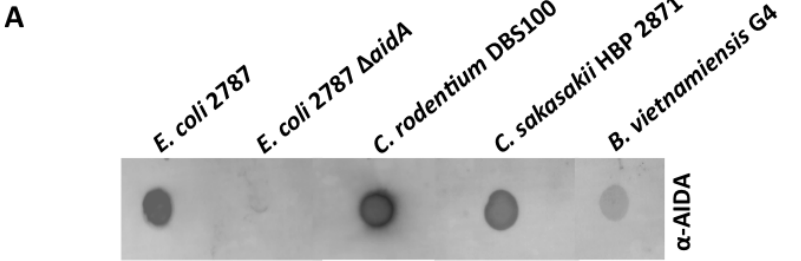


Figure 3

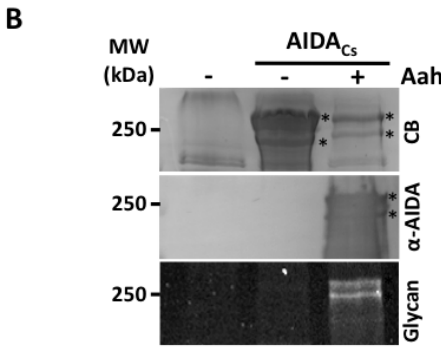
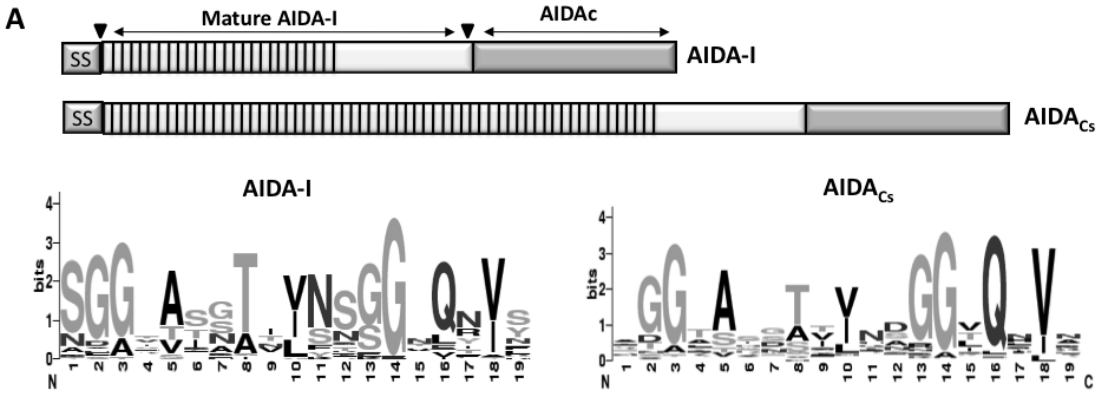


Figure 4

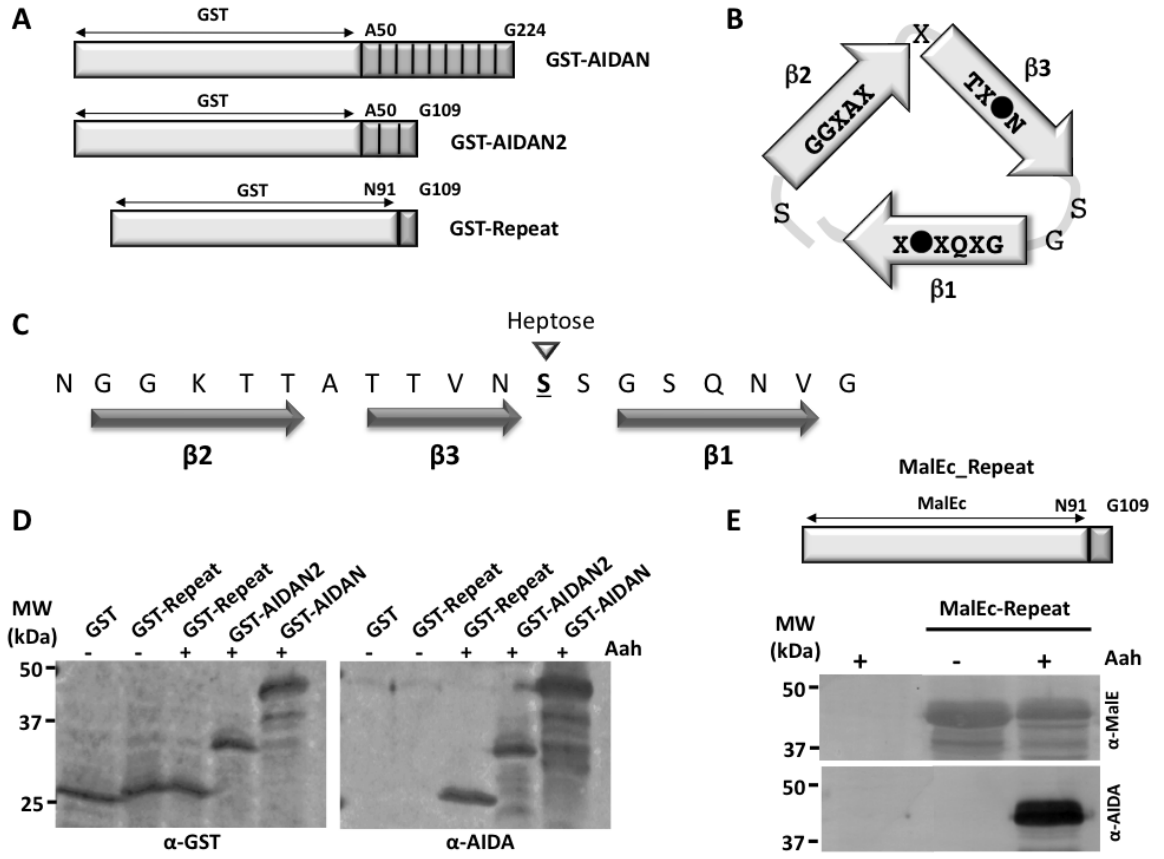


Figure 5

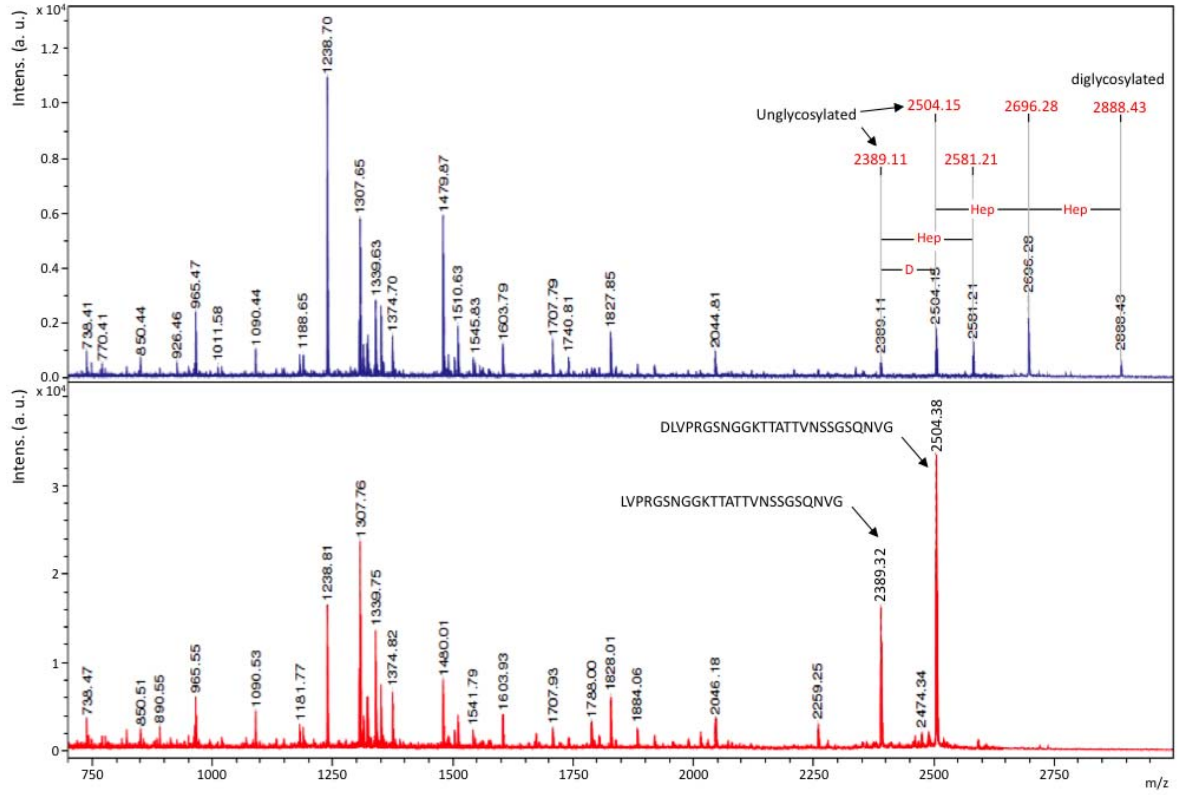


Figure 6

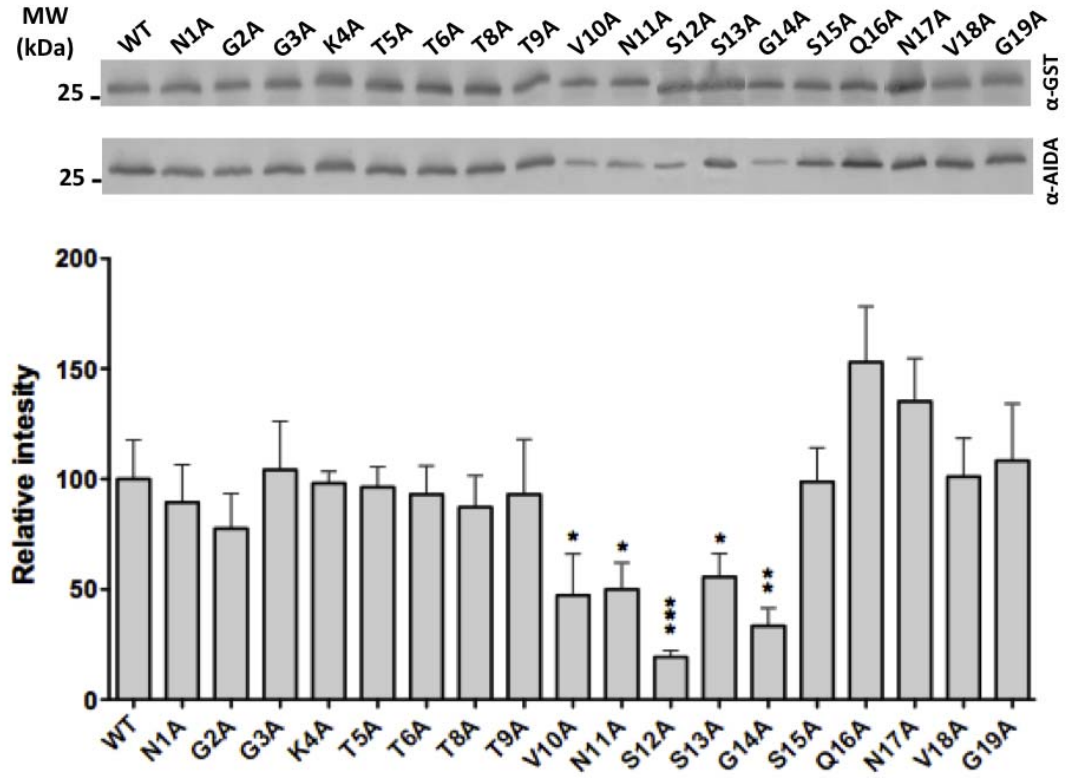




Figure 7

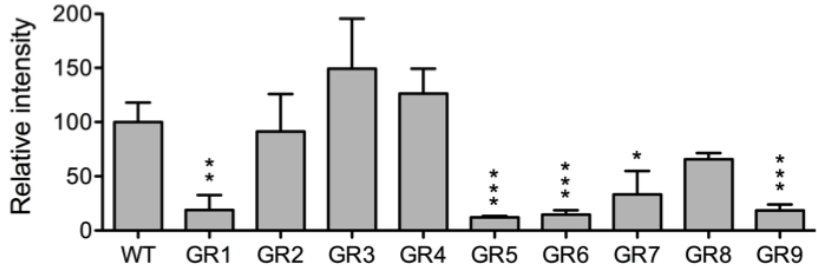
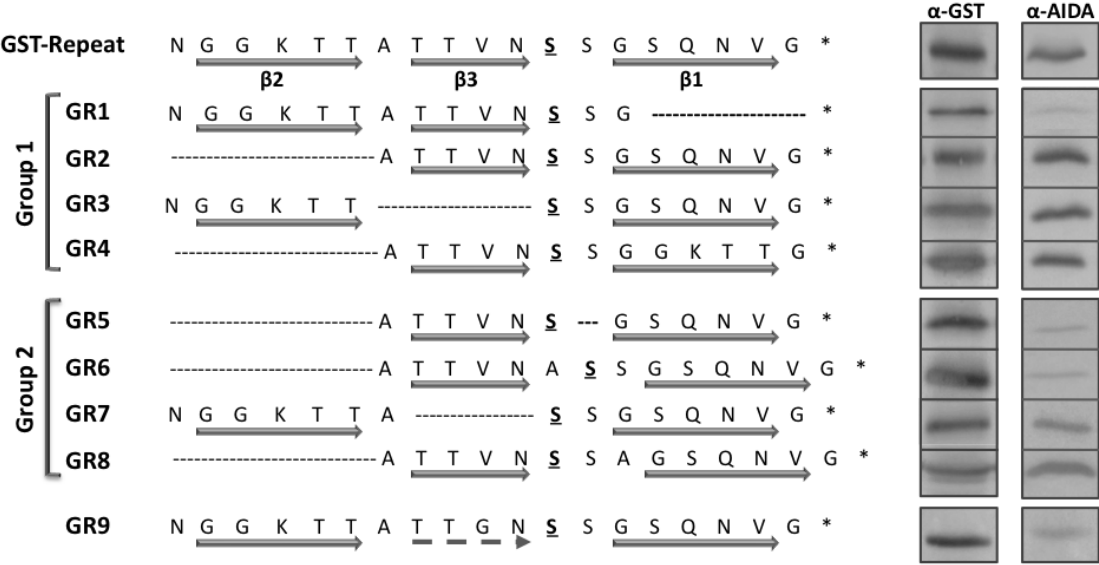
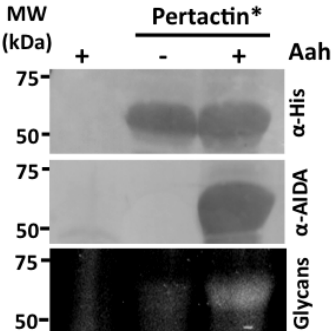


Figure 8



## Supporting information

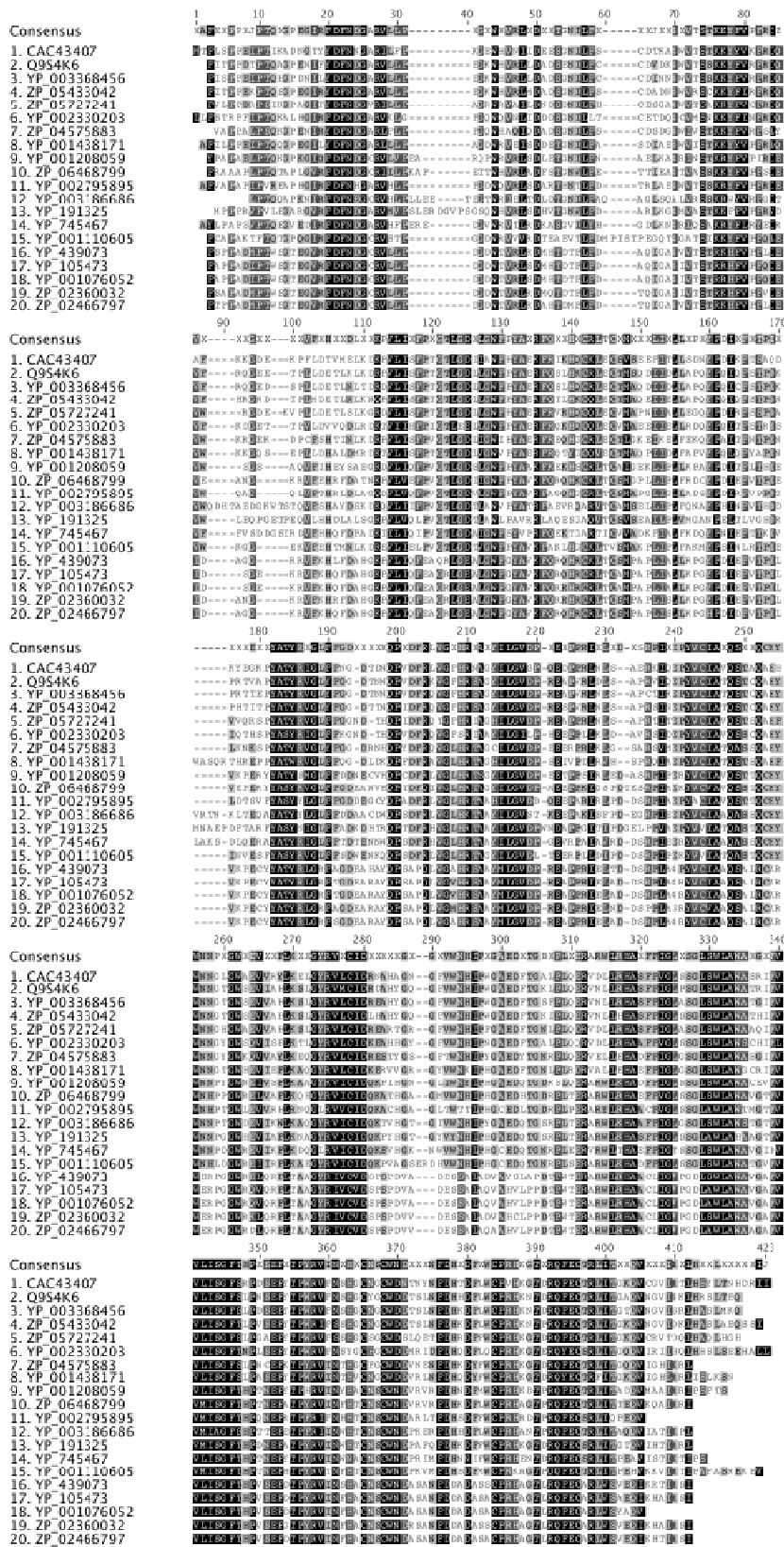
Table SI: Primers used in this study

Primer	Sequence
AidaN_109	5'-GGAAGCCAGAACGTCGGGTA ACTCGAGCGGCCGCATCG-3'
pGEX_Repeat	5'-GGATCTGGTTCCGCGTGGATCCAATGGTGGGAAAACCAC TG-3'
Repeat_N1A	5'-CGGATCTGGTTCCGCGTGGCTCCGCTGGTGGGAAAACCA CTGC-3'
Repeat_G2A	5'-CGGATCTGGTTCCGCGTGGCTCCAATGCTGGGAAAACCA CTGC-3'
Repeat_G3A	5'-GGATCTGGTTCCGCGTGGCTCCAATGGTGGGAAAACCAC TGCTACA ACTG-3'
Repeat_K4A	5'-GGATCTGGTTCCGCGTGGATCAAATGGTGGGGCAACCA CTGCTACA ACTG-3'
Repeat_T5A	5'-GGTTCGCGTGGATCAAATGGTGGGAAAGCCACTGCTAC AACTG-3'
Repeat_T6A	5'-GGTGGGAAAACCGCTGCTACA ACTGTTAACAGTTCAGG AAGCCAGAACG-3'
Repeat_T8A	5'-GGGAAAACC ACTGCTGCAACTGTTAACAGTTCAGGAAG CCAGAACG-3'
Repeat_T9A	5'-GGAAAACC ACTGCTACAGCTGTTAACAGTTCAGGAAG CCAGAACG-3'
Repeat_V10A	5'-CCACTGCTACA ACTGCGAATTCTTCAGGAAGCCAGAAC G-3'
Repeat_N11A	5'-CCACTGCTACA ACTGTTGCTAGCTCAGGAAGCCAGAAC G-3'
Repeat_S12A	5'-GCTACA ACTGTTAACGCTTCAGGAAGCCAGAACG-3'
Repeat_S13A	5'-GCTACA ACTGTTAACAGTGCAGGAAGCCAGAACGTCGG-3'
Repeat_S12/13A	5'-CTGCTACA ACTGTTAACGCTGCAGGAAGCCAGAACGTC GG-3'
Repeat_G14A	5'-GCTACA ACTGTTAATAGTTCAGCTAGCCAGAACGTCGG G-3'
Repeat_S15A	5'-GCTACA ACTGTTAATAGTTCAGGTGCACAGAACGTCGG GTA ACTCG-3'
Repeat_Q16A	5'-CTGTTAATAGTTCAGGATCCGCGAACGTCGGGTA ACT CG-3'
Repeat_N17A	5'-CTGTTAATAGTTCAGGATCCCAGGCCGTCGGGTA ACTC G-3'
Repeat_V18A	5'-CAGGAAGCCAGAACGCCGGTTA ACTCGAGCGGCCG-3'
Repeat_G19A	5'-GGAAGCCAGAACGTCGCGTAACTGGAGCGGCCGCATCG-3'
GR1	5'-GCTACA ACTGTTAATAGTTCATA ACTAGAGCGGCCGCAT CGTG-3'

GR2	5'-CGGATCTGGTTCCGCGTGGATCGGCTACAACGTGTTAATAGTTCAGG-3'
GR3	5'-GGTGGGAAAACCACTAGTTCAGGAAGCCAGAACG-3'
GR4	5'-CTGTTAATAGTTCAGGAGGCAAGACCACCGCGTAATTCGAGCGGGAAAGCCAGAACG-3'
GR5	5'-CGGCTACAACGTGTTAACAGTGGAAGCCAGAACGTCGGG-3'
GR6	5'-CGGCTACAACGTGTTAACGCAAGTTCAGGAAGCCAGAACG-3'
GR7	5'-GGTGGGAAAACCACTGCTAGCTCAGGAAGCCAGAACGTCGGG-3'
GR8	5'-CTGTTATAGTTCAGCCGGCAGCCAGAACGTCGGG-3'
GR9	5'-CCACTGCTACAACCGGTAATAGTTCAGGAAGC-3'
MalE_Repeat_F	5'-ATATATGAATTC AATGGTGGGAAAACCACTGC-3'
MalE_Repeat_R	5'-ATATATTCTAGATTACCCGACGTTCTGGCTTCC-3'
Pertactin_F	5'-ATATATCCATGGACTGGAACAACCAGTCTATCG-3'
Pertactin_R	5'-ATATATTTCTAGATTAACCAACCAGAGACCACTGACC-3'
pPertHis	5'-GGAAACAGACCATGCATCATCATCATCACGACTGGAACAACCAG-3'
AIDA <sub>Cs</sub> _F	5'-TATATCCATGGGAAAGAGAAGTCTTAATTGCAGTTACCGGC-3'
AIDA <sub>Cs</sub> _F2	5'-ATAATCCCGGGTCGAGAAAGGCGGCGTTC-3'
AIDA <sub>Cs</sub> _R	5'-ATAATCCCGGGTGTGCAAGGCTTTACCGC-3'
AIDA <sub>Cs</sub> _R2	5'-TATAATCTAGATCAGAAGCTGACTTTCAGCGTCAC-3'
Aah <sub>CR</sub> _F	5'-TATAACATGTTGAAGTCGTTAGCAACGTTTTTTATTCTCCTCCG-3'
Aah <sub>CR</sub> _R	5'-TATTTCTCGAGCCCGGTGGATTTCATTCCATAC-3'
AIDA <sub>Cr</sub> _KO_F	5'-CCTGATGAAGCAGGGAGATAAAGCCTGTCTTACAAAAGGAAAATGATTCCGGGGATCCGTCGACC-3'
AIDA <sub>Cr</sub> _KO_R	5'-ACCTGATGTCCGTCCCTCTTTTCCCTACATCAGAAATTAATCCGGAAACCTGTAGGCTGGAGCTGCTTCG-3'
Aah <sub>Cr</sub> _KO_F	5'-TAATTATCCTACACTTCATTTAATAATATTATTCTGGTGATAAAAGTATGATTCCGGGGATCCGTCGACC-3'
Aah <sub>Cr</sub> _KO_R	5'-ACATCCCGGTGGATTTCATTCCATACAGTGTATAAACCTTGTTTCATTGTTTGTAGGCTGGAGCTGCTTCG-3'
Exc_AIDA_F	5'-ACCACAGTGAATGTTTCTGGTAC-3'
Exc_AIDA_R	5'-CGTTTCCTTTCTGCAGTGTGTAATC-3'

---

Fig. S1 :



**Fig. S1 : Alignment of 20 Aah homologs**

Sequence alignment of 20 Aah homologs was performed using the BLOSUM 62 matrix, with the Geneious Pro 4.0.2 software (Biomatters Ltd.).

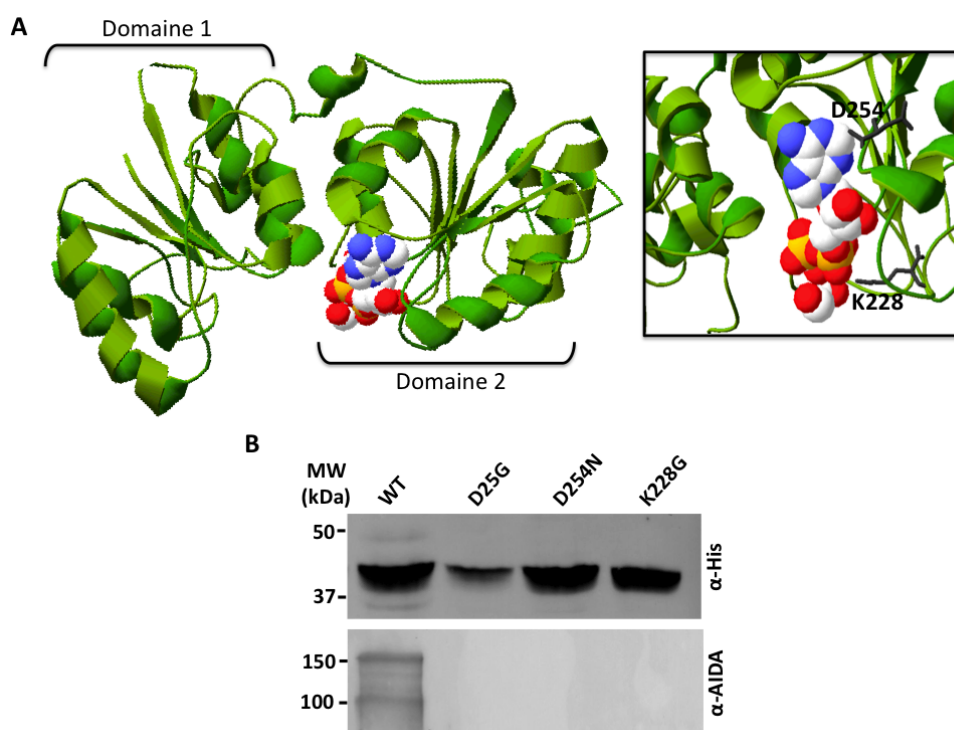
## Résultats supplémentaires

Les résultats présentés dans cette section ne sont pas encore organisés sous la forme d'une publication scientifique. Toutefois, un article regroupant la plupart des résultats présentés ici est en cours de rédaction et s'intitulera « Characterization of a new family of *O*-glycosyltransferases uncovers an unique cotranslational mechanism of glycosylation in bacteria ». J'ai effectué toutes les expérimentations présentées dans cette section, à l'exception du modèle en trois dimensions de Aah.

L'objectif derrière les expérimentations décrites ci-dessous est de caractériser la nouvelle famille de glycosyltransférases décrites dans l'article 6 du présent manuscrit en utilisant Aah à titre de prototype. En effet, Aah et ses homologues, dont TibC (la glycosyltransférase associée à l'autotransporteur TibA), possèdent un mécanisme d'action unique. Aah permet la glycosylation de son substrat, AIDA-I, dans le cytoplasme, et ce, indépendamment du transport à travers la membrane interne. Ce mode d'action contraste avec tous les autres systèmes de *O*-glycosylation décrits précédemment chez les bactéries Gram négatif, où la glycosylation a lieu dans le périplasme. Cette glycosylation cytoplasmique implique nécessairement un mécanisme de régulation permettant le contrôle et la coordination des événements de traduction, glycosylation et sécrétion, afin que le polypeptide puisse être complètement glycosylé avant d'être transporté dans le périplasme.

## 1) Détermination du site catalytique de la glycosyltransférase Aah

Aah permet la glycosylation de AIDA-I avec des résidus heptose empruntés de la voie de biosynthèse des LPS. Il est donc possible que le site catalytique de Aah ressemble à celui des heptosyltransférases impliquées dans la biosynthèse des LPS chez *E. coli*, soit RfaF (WaaF), RfaQ (WaaQ) et RfaC (WaaC) (189). Présentement, les structures de RfaF et WaaC sont disponibles (362). En utilisant un alignement multiple entre Aah et les heptosyltransférases RfaF et WaaC, Dr. Peter Pawelek (Université Concordia) a généré un modèle en trois dimensions de Aah (Fig. 1A). Le modèle obtenu pour Aah montre une structure typique du groupe de glycosyltransférases possédant un repliement de type GT-B, caractérisé par deux domaines avec un motif Rossmann ( $\beta$ - $\alpha$ - $\beta$ ) séparés par une région de jonction (363).



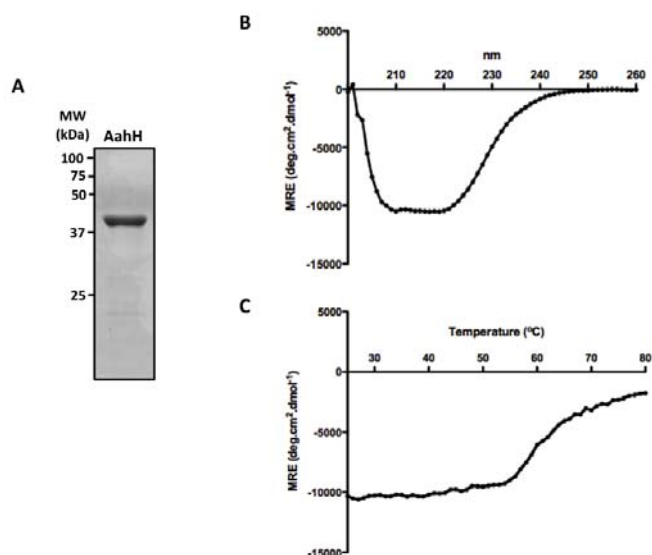
**Figure 1 : Modèle structural de Aah représentant les résidus importants pour la fonctionnalité. A.** Modèle en trois dimensions de Aah basé sur un alignement avec les heptosyltransférases WaaC et RfaF de *E. coli*. La molécule d'ADP-2F-heptose, un analogue du sucre donneur, co-cristallisée avec WaaC est aussi représentée et se situe dans le creux formé par les deux domaines possédant un repliement de type Rossmann. Le grossissement du site d'interaction entre Aah et le sucre est présenté dans l'encadré. La localisation de deux résidus de Aah (D254 et K228), importants pour la fonctionnalité, est indiquée dans le modèle structural. **B.** Extraits totaux de la souche *E. coli* BL21, exprimant Aah sauvage (WT) ou un des mutants ponctuels de Aah, faits à partir de cultures induites avec 10  $\mu$ M IPTG à DO<sub>600nm</sub> de 0,7 pour la nuit. Les protéines séparées par SDS-PAGE ont été révélées par un anticorps contre l'étiquette six histidines ou contre AIDA-I mature glycosylé.



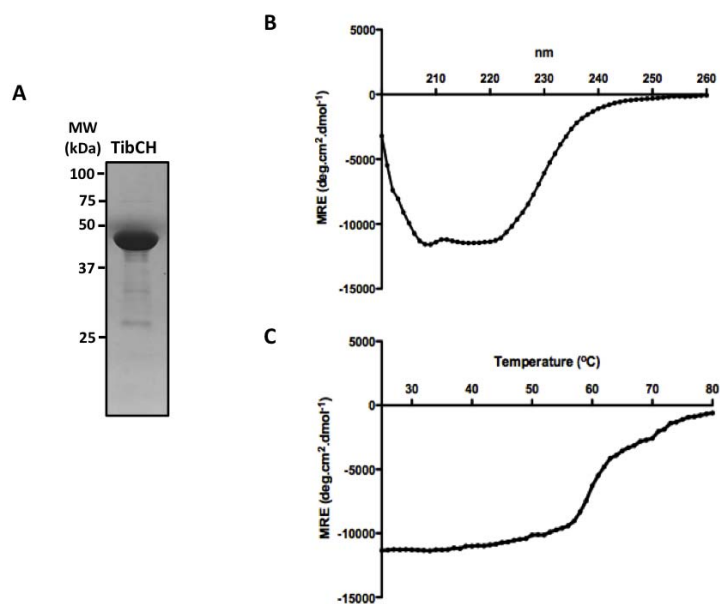
Le mécanisme catalytique de WaaC est connu et implique le motif D(S/T)(G/A)XXH nommé « signature heptosyltransférase » ainsi que trois résidus bien conservés, soit une aspartate en position 13, une lysine en position 192 et un glutamate en position 222 (362). En superposant le modèle de Aah avec la structure de WaaC, on a réussi à identifier un aspartate (position 25), une lysine (position 228) et un glutamate (position 254) dans Aah qui se positionnent de façon similaire à ceux retrouvés chez WaaC (Fig. 1A). Pour valider l'importance de ces résidus, nous avons effectué des mutations ponctuelles qui modifient le groupement fonctionnel de ces acides aminés spécifiques. Les résultats présentés dans la figure 1B montrent bien que ces trois résidus sont essentiels pour le fonctionnement de Aah, puisque les trois protéines résultantes sont incapables de glycosyler AIDA-I. Il faut toutefois noter que la mutation D25G semble aussi affecter la stabilité de Aah. Le motif heptosyltransférase de six acides aminés n'est pas retrouvé chez Aah suggérant que des différences majeures entre Aah et WaaC existent au niveau du mécanisme d'action.

## 2) Purification de Aah et TibC

Nous avons fait une caractérisation biochimique de ces deux glycosyltransférases à partir des protéines purifiées. Une étiquette six histidines a été insérée en C-terminal pour permettre la purification. Les figures 2 (AahH) et 3 (TibCH) montrent les résultats des purifications. Les profils obtenus par dichroïsme circulaire (CD) ont été analysés par le logiciel Dichroweb ([www.cryst.bbk.ac.uk/cdweb](http://www.cryst.bbk.ac.uk/cdweb)) afin de déterminer la composition en élément de structure secondaire pour ces deux protéines. Aah est composé à 28% d'hélices  $\alpha$  et 25% de feuillets  $\beta$ , résultats similaires à ceux obtenus pour TibC, soit 26% d'hélices  $\alpha$  et 15% de feuillets  $\beta$ . De plus, leur température de fusion ( $T_m$ ) lors de la dénaturation thermique est assez similaire, soit un  $T_m$  de  $58.95 \pm 0.14^\circ\text{C}$  pour Aah et  $59.12 \pm 0.12^\circ\text{C}$  pour TibC. Ces résultats nous indiquent que ces deux glycosyltransférases ont une structure assez semblable. De plus, comme les résultats obtenus avec le CD concordent avec la composition en structure secondaire calculée à partir du modèle de Aah (31% d'hélices  $\alpha$  et 14% de feuillets  $\beta$ ), on peut donc considérer que ces deux protéines sont repliées, possiblement dans une conformation native.



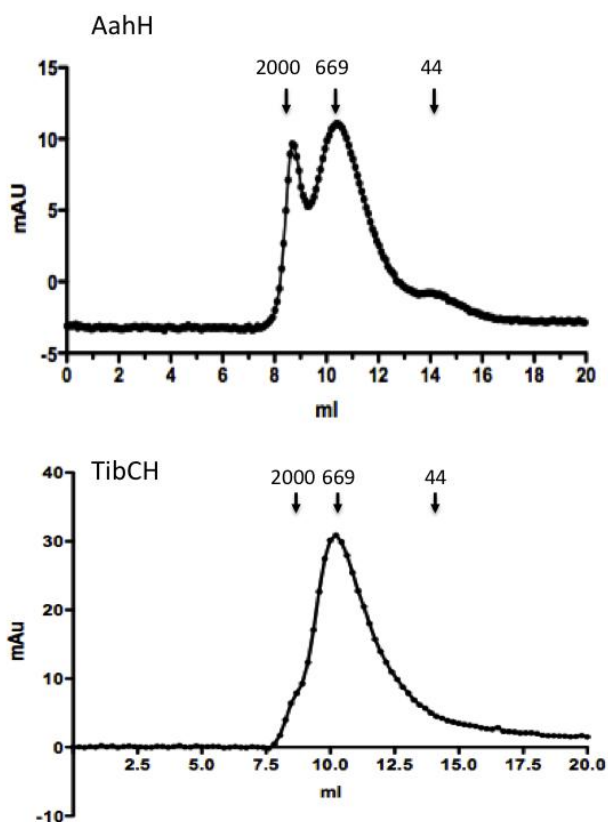
**Figure 2 : Purification et analyse de AahH par CD.** A. Purification de AahH à partir d'un litre de *E. coli* BL21 possédant le plasmide pAah, induit pour 3 heures à une  $DO_{600nm}$  de 0.4 avec 10  $\mu$ M d'IPTG. Après la lyse, la fraction soluble a été isolée par une centrifugation de 45 minutes à 16 000 x g. AahH a été purifié en 50 mM NaCl, 50 mM Tris pH8 et 0.005% Triton X-100 réduit avec une colonne HisTrap de 1 ml sur un système de purification AKTA. La pureté de la protéine a été vérifiée sur gel SDS-PAGE coloré au bleu de Coomassie. B. Cette figure montre le spectre de CD résultant de la compilation de 10 accumulations fait sur 200  $\mu$ g de protéines. C. Dénaturation thermique effectuée entre 30 et 70°C sur 200  $\mu$ g de protéines. La figure montre l'ellipticité à 218 nm. MRE; ellipticité moyenne résiduelle.



**Figure 3 : Purification et analyse de TibCH par CD.** A. Purification de TibCH à partir d'un litre de *E. coli* BL21 possédant le plasmide pTibCH, induit pour 3 heures à une  $DO_{600nm}$  de 0.4 avec 10  $\mu$ M d'IPTG. Après la lyse, la fraction soluble a été isolée par une centrifugation de 45 minutes à 16 000 x g. TibCH a été purifié en 50 mM NaCl, 50 mM Tris pH8 et 0.1% Triton X-100 réduit avec une colonne HisTrap de 1 ml sur un système de purification AKTA. La pureté de la protéine a été vérifiée sur gel SDS-PAGE coloré au bleu de Coomassie. B. Cette figure montre le spectre de CD résultant de la compilation de 10 accumulations fait sur 200  $\mu$ g de protéines. C. Dénaturation thermique effectuée entre 30 et 70°C sur 200  $\mu$ g de protéines. La figure montre l'ellipticité à 218 nm. MRE; ellipticité moyenne résiduelle.

### 3) Oligomérisation de Aah et TibC

Dans le but de mieux comprendre le mode de fonctionnement de ces deux protéines, nous avons vérifié leur état d'oligomérisation en passant les protéines purifiées sur une colonne de gel filtration, soit une colonne de séphacryl S-300 haute résolution de 20 ml. Les résultats présentés dans la figure 4 montrent clairement que ces deux glycosyltransférases forment des oligomères de haut poids moléculaire, soit d'environ 600 kDa. Considérant que les poids prédits pour Aah et TibC sont respectivement de 45 kDa et 46 kDa, des oligomères d'environ 600 kDa impliquent l'assemblage d'un grand nombre de sous-unités. Cette capacité d'oligomérisation est exceptionnelle pour une heptosyltransférase puisque celles impliquées dans la biosynthèse des LPS sont monomériques (362), illustrant encore les différences majeures qui existent entre ces deux familles de protéines.

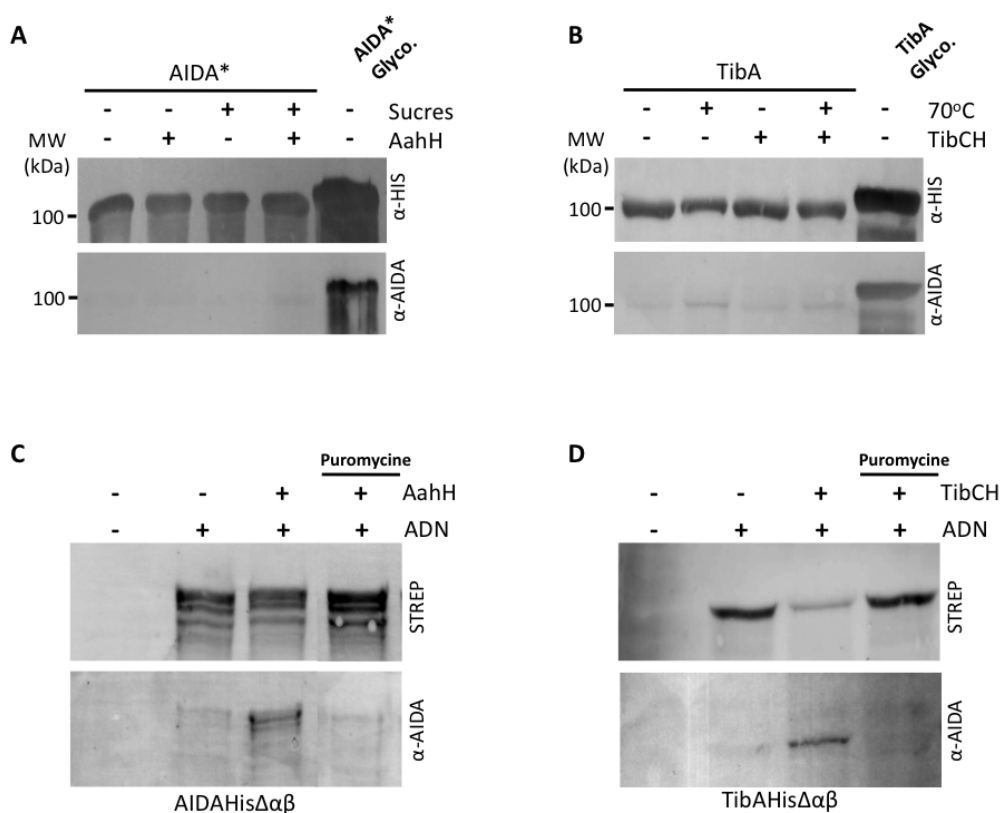


**Figure 4 : Oligomérisation de AahH et TibCH.** Les profils de gel filtration pour AahH et TibCH résultant d'un chargement de 150 $\mu$ g de protéines sur une colonne de 20 ml de séphacryl S-300 haute résolution. Les poids moléculaires en kDa de trois protéines standard sont indiqués sur les graphiques.

#### 4) Glycosylation co-traductionnelle

Maintenant que nous possédons les deux glycosyltransférases purifiées, nous avons voulu reproduire la réaction de glycosylation dans un système *in vitro*. Toutefois, l'instauration d'un tel système est plutôt complexe, principalement du au fait que les heptoses (sucres ajoutés sur AIDA-I et TibA) ne sont pas disponibles commercialement. Pour obtenir une source de sucres, nous avons utilisé une procédure décrite précédemment qui consiste à isoler par précipitation à l'éthanol et centrifugation avec un filtre de 3 kDa, des molécules de faibles poids moléculaire à partir d'une culture bactérienne en phase exponentielle de croissance (364). La question du substrat utilisé pour faire le test de glycosylation *in vitro* est aussi problématique. En effet, AIDA-I ne peut pas être purifié sous sa forme non-glycosylée puisque la glycosylation est requise pour la stabilité (voir article 3). Pour cette raison, nous avons utilisé une construction décrite dans l'article 5 et nommée pAngH $\Delta$ SS $\Delta$  $\alpha\beta$ . Cette construction permet l'expression de AIDA-I sans séquence signal, sans tonneau  $\beta$  et sans l'hélice  $\alpha$  (AIDA\*). La purification de cette protéine est donc effectuée à partir du cytoplasme de la bactérie. Pour TibC, nous avons utilisé comme substrat du TibA non-glycosylé purifié à partir de la membrane externe (la glycosylation n'est pas aussi importante pour la stabilité de TibA que pour celle de AIDA-I, Jean-Philippe Côté et Michael Mourez, manuscrit en préparation). Le test de glycosylation *in vitro* a été fait en incubant le substrat et la glycosyltransférase dans un ratio de 1 :1 pendant 2 heures à 37°C, puis toute la nuit à 4°C. La glycosylation est détectée par la révélation de AIDA-I avec l'anticorps dirigé spécifiquement contre la partie AIDA-I mature glycosylée. Malgré nos nombreux essais, la glycosylation dans ce type de système *in vitro* ne fonctionne pas (Figure 5A, B).

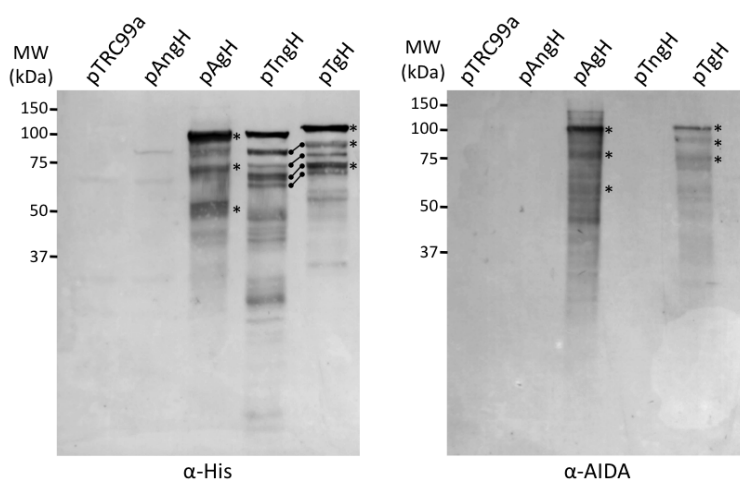
Afin d'optimiser ce système de glycosylation *in vitro*, nous avons refait cette expérience en utilisant un système de transcription/traduction *in vitro* (kit PURExpress, New England Biolabs). Dans ce système, le substrat des glycosyltransférases n'est pas une protéine purifiée mais une protéine nouvellement traduite ou en cours de traduction. Les gènes codant pour AIDA-I et TibA, sans domaine  $\beta$ , ont donc été clonés sous le contrôle d'un promoteur T7 (plasmide pIVEX 2.3-MCS). Afin de détecter les protéines nouvellement synthétisées, des ARNt pour la lysine, marqués à la biotine, ont été ajoutés à la réaction, permettant ainsi une détection par la streptavidine couplée à la HRP (Transcend™ Non-Radioactive Translation Detection Systems, Promega).



**Figure 5 : Glycosylation *in vitro*.** **A.** AIDA\*, la source de sucres ainsi que AahH ont été incubés pour deux heures à 37°C, suivi d'une incubation sur la nuit à 4°C. Les protéines séparées par SDS-PAGE ont été révélées à l'aide d'un anticorps contre l'étiquette six histidines ou contre AIDA-I mature glycosylé. AIDA\* glycosylé a été utilisé comme contrôle. **B.** Procédure similaire à A, mais avec TibA non-glycosylé comme substrat. Comme TibA a été purifié sous sa forme native, un traitement de 10 minutes à 70°C a été fait pour dénaturer la protéine. TibA natif glycosylé a été utilisé comme contrôle. **C.** Glycosylation dans un système de transcription/traduction *in vitro*. Les composants du kit, Aah, les sucres et l'ADN ont été incubés 90 minutes à 37°C avec d'être déposés sur un gel SDS-PAGE. Pour la ligne puromycine, la réaction de traduction a eu lieu pendant 90 minutes pour permettre la synthèse de nouvelles protéines avant l'ajout de la puromycine, de Aah et des sucres. La réaction s'est poursuivie pour un autre 90 minutes à 37°C. **D.** Procédure similaire à C, mais avec le gène codant pour TibAHisΔαβ comme ADN substrat.

Comme illustré à la figure 5C et D, le test est fonctionnel dans ces conditions, ce qui signifie que Aah et TibC purifiés peuvent ajouter des sucres sur AIDA-I et TibA, respectivement. Ce résultat suggère donc que ces deux glycosyltransférases modifient leur substrat uniquement lors du processus de traduction. Pour vérifier cette hypothèse, nous avons refait cette réaction en permettant une synthèse initiale des protéines pendant 90 minutes, puis de la puromycine, un inhibiteur de la traduction, a été ajoutée. La glycosyltransférases et la source de sucres ont par la suite été ajoutées, suivi par une seconde incubation de 90 minutes. Dans ces conditions, il n'y pas de glycosylation, suggérant clairement que la glycosylation est un processus co-traductionnel.

Afin de confirmer ces observations *in vivo*, nous avons décidé d'isoler des intermédiaires de traduction. L'intérêt est de voir si les chaînes polypeptides toujours associées aux ribosomes sont déjà glycosylées. Pour ce faire, nous avons utilisé les plasmides permettant l'expression de AIDA-I non-glycosylé (pAngH), AIDA-I glycosylé (pAgH), TibA non-glycosylé (pTngH) ou TibA glycosylé (pTgH). Les cultures ont été incubées à 30°C en agitation jusqu'à une  $DO_{600nm}$  de 0.3 avant d'être transférées pour 30 minutes à 20°C, la baisse de température favorisant un ralentissement de la traduction. La synthèse des protéines a été induite par l'ajout de 100  $\mu$ M d'IPTG pendant 15 minutes. Afin de bloquer complètement la traduction, les cultures ont été incubées à 20°C en présence de 50  $\mu$ g/ml de chloramphénicol pour 5 minutes.



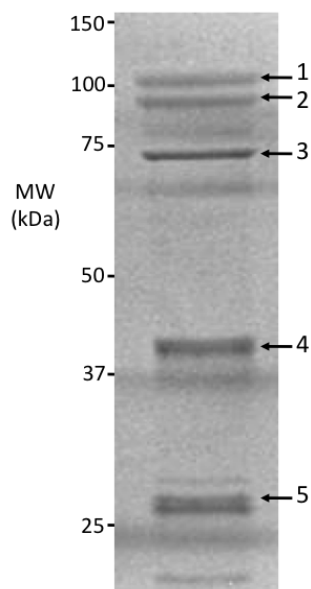
**Figure 6 : Isolation d'intermédiaires de traduction glycosylés.** Des cultures de *E. coli* BL21 possédant le plasmide pTRC99a, pAngH, pAgH, pTngH ou pTgH ont été incubées à 30°C jusqu'à l'obtention d'une  $OD_{600nm}$  de 0,3 avant d'être transférées pour 30 minutes à 20°C. Après une incubation de 15 minutes avec 100  $\mu$ M d'IPTG, du chloramphénicol à 50  $\mu$ g/ml a été ajouté pour 5 minutes, suivi par la lyse des cellules et le dépôt sur gel SDS-PAGE. La révélation des protéines a été faite à l'aide d'un anticorps contre l'étiquette six histidines ou contre AIDA-I mature glycosylé.

La figure 6 montre un résultat caractéristique. Comme toutes ces protéines possèdent une étiquette six histidines en N-terminal, le profil de bandes visibles avec l'anticorps dirigé contre cette étiquette correspond fort probablement à des protéines en cours de traduction. Dans le cas de AIDA-I, aucune conclusion claire ne peut être émise puisque le AIDA-I non-glycosylé est à peine visible, conséquence du problème de stabilité de cette protéine. Par contre, pour TibA, le profil de bandes entre les formes non-glycosylée et glycosylée est similaire, à l'exception de la taille. La mobilité aberrante sur gel est généralement causée

par la glycosylation de la protéine, observation qui est confirmée par la détection de ces mêmes bandes avec l'anticorps contre AIDA-I mature glycosylé (les bandes sont indiquées par des astérisques sur la figure 6). Ce résultat indique donc que les intermédiaires de traduction isolés à partir des bactéries exprimant TibA et la glycosyltransférase TibC sont déjà modifiés par l'ajout de sucres.

### 5) Co-purification de composants cytoplasmiques avec AIDA\*

La protéine nommée AIDA\* correspond à une version tronquée de AIDA-I, c'est-à-dire une protéine sans tonneau  $\beta$  et d'hélice  $\alpha$  qui est exprimée uniquement dans le cytoplasme de la bactérie due à l'absence de séquence signal (voir l'article 5 pour une description complète). À haut niveau d'induction, soit 1 mM d'IPTG, cette protéine forme des corps d'inclusion dans le cytoplasme. Toutefois, avec une induction à 10  $\mu$ M d'IPTG seulement, elle peut être purifiée sous forme soluble à partir du cytoplasme grâce à une étiquette six histidines localisée en N-terminale. Nous avons effectué cette purification pour constater que plusieurs autres protéines étaient co-purifiées avec AIDA-I\*.



**Figure 7 : Co-purification de composants cytoplasmiques avec AIDA\*.** Purification de AIDA\* à partir d'un litre de *E. coli* BL21 possédant le plasmide pAngH $\Delta$ SS $\Delta$  $\alpha$  $\beta$ , induit pour 3 heures à une  $DO_{600nm}$  de 0.4 avec 10  $\mu$ M d'IPTG. Après la lyse, la fraction soluble a été isolée par une centrifugation de 45 minutes à 16 000 x g. AIDA\* a été purifié avec une colonne HisTrap de 1 ml sur un système de purification AKTA. La protéine purifiée a été vérifiée sur gel SDS-PAGE coloré au bleu de Coomassie.

Les cinq bandes majeures qui apparaissent sur le gel présenté à la figure 7 ont été identifiées par spectrométrie de masse. Les bandes 1 et 2 correspondent au pro-AIDA\* avant clivage (90 kDa) et au AIDA\* mature (79 kDa), respectivement. La bande 3 a été identifiée comme étant la chaperonne cytoplasmique DnaK, alors que la bande 4 est la glycosyltransférase Aah. Finalement, la cinquième bande a été identifiée comme la protéine ribosomale S2 du complexe S30. Ces résultats sont présentés ici, puisqu'il est quand même intéressant de constater qu'on peut co-purifier un complexe contenant AIDA-I, Aah, une chaperonne cytoplasmique ainsi qu'une sous-unité ribosomale, surtout considérant que la glycosylation d'AIDA-I est un processus co-traductionnel.



## **Discussion générale**

Les autotransporteurs monomériques sont une famille importante de facteurs de virulence bactériens. De plus, ils appartiennent au système de sécrétion de type V, souvent considéré comme le système de sécrétion le plus simple et le plus répandu chez les bactéries Gram négatif. Plusieurs fonctions, souvent essentielles pour le développement d'une infection ou pour le maintien et la survie des bactéries dans l'organisme hôte, telles que l'adhésion et l'invasion cellulaire, la résistance au complément, une activité cytotoxique, la mobilité intracellulaire, parmi d'autres, ont été décrites pour la famille des autotransporteurs. Une meilleure connaissance de ces protéines autotransportrices est donc requise afin de mieux comprendre la pathogenèse reliée à de nombreux micro-organismes bactériens. De plus, puisque les autotransporteurs monomériques sont de plus en plus utilisés pour des applications biotechnologiques, une meilleure compréhension de leur biogenèse est essentielle.

Dans cette discussion, les différentes étapes de la biogenèse d'un autotransporteur monomérique typique, soit l'adhésine impliquée dans l'adhérence diffuse (AIDA-I) retrouvé chez *E. coli*, vont être discutées en détail afin de mettre en évidence les avancées présentées dans cette thèse de doctorat.

### **1) Glycosylation cytoplasmique**

La glycosylation est une des modifications les plus répandues chez les eucaryotes. Toutefois, L'étude de la glycosylation de protéines bactériennes est un champ d'investigation relativement récent. En fait, on commence à peine à comprendre l'importance de cette modification pour la physiologie bactérienne ainsi que pour la virulence. Les travaux présentés dans cette thèse ont permis de définir une nouvelle famille de glycosyltransférases bactériennes possédant un mécanisme d'action unique et qui sont importantes pour la fonction de plusieurs protéines considérées comme des facteurs de virulence.

Nous avons tout d'abord voulu étudier le mécanisme d'action de Aah, une des premières glycosyltransférases identifiées chez *E. coli*. Les glycosyltransférases sont principalement

classifiées selon leur capacité à ajouter des sucres sur des résidus asparagine (N-glycosylation) ou sérine et thréonine (O-glycosylation). En utilisant la spectrométrie de masse et le séquençage par dégradation d'Edman, nous avons pu démontrer qu'AIDA-I est une protéine O-glycosylée (article 3). Plus précisément, 15 sérines et une thréonine sont modifiées par des résidus heptose, avec une possibilité de plus de 19 sucres, ce qui concorde avec une étude antérieure qui suggérait un ratio de 19 heptoses par molécule d'AIDA-I (7). AIDA-I est donc le premier autotransporteur pour lequel une O-glycosylation a été démontrée expérimentalement.

La plupart des glycosyltransférases bactériennes sont retrouvées dans des locus génétiques comportant toutes les enzymes requises pour la synthèse du sucre, son transfert dans le périplasma ainsi que pour l'ajout de ce sucre sur une protéine cible (192). Une des différences majeures entre ces systèmes de O-glycosylation et Aah est l'absence de ce type de locus génétique. En effet, le gène codant pour Aah est retrouvé en opéron avec le gène codant pour son substrat AIDA-I uniquement (7), sans autres facteurs accessoires, suggérant un mécanisme d'action différent. En accord avec cette hypothèse, nos résultats montrent clairement que AIDA-I est glycosylé dans le cytoplasme, indépendamment de son transport à travers la membrane interne. En effet, l'expression d'AIDA-I sans séquence signal ou le blocage de la sécrétion sec-dépendante par la croissance des bactéries en présence de sodium azide n'affecte pas le niveau de glycosylation d'AIDA-I (article 4). Cette observation est importante puisqu'une glycosylation cytoplasmique est assez unique chez les bactéries Gram négatif. Pour l'instant, un seul autre exemple est disponible dans la littérature et correspond à la N-glycosyltransférase HMW1C (et ses homologues) (207). La plupart des systèmes de glycosylation permettent l'ajout de sucres sur les protéines dans le périplasma, ce qui est nécessaire afin d'éviter de bloquer la machinerie de sécrétion Sec par la présence de sucres complexes sur les polypeptides à sécréter. Toutefois, dans le cas de Aah et HMW1C, les sucres ajoutés sont des sucres simples, soit un heptose (7) et un glucose, galactose ou maltose (207), respectivement, suggérant ainsi que la machinerie Sec peut facilement accommoder des polypeptides avec des modifications mineurs. Il faut néanmoins mentionner que la glycosylation de plusieurs protéines appartenant à la famille des adhésines riches en sérine chez les bactéries Gram positif a lieu dans le cytoplasme (246). Toutefois, dans ces cas bien particuliers, les protéines modifiées sont sécrétées par un système de sécrétion Sec accessoire, différents du système Sec classique, et qui, *a priori*,

formerait un canal d'une taille suffisante pour accommoder des polypeptides hautement modifiés. Toutefois, un tel système Sec accessoire n'est pas retrouvé chez les bactéries Gram négatif.

Une étude plus détaillée de la glycosyltransférase Aah nous a permis de confirmer la glycosylation cytoplasmique d'AIDA-I. En effet, Aah est une protéine sans séquence signal ni domaine transmembranaire qui se localise dans le cytoplasme. Par contre, en regardant la localisation précise de cette enzyme, nous avons observé qu'une fraction relativement importante de Aah se retrouvait aussi associée à la membrane interne (article 4). L'observation d'une association avec la membrane interne n'est pas surprenante puisque les heptosyltransférases impliquées dans la biosynthèse des LPS chez *E. coli*, qui présentent un certain niveau de similarité avec Aah, sont des protéines périphériques de membrane (364). Il est donc possible que les sucres utilisés par Aah soient aussi localisés dans l'environnement périphérique de la membrane interne. Il faut mentionner qu'une partie non négligeable de Aah se retrouve dans les agrégats ou débris cellulaires, suggérant que la protéine a une tendance à agréger, tendance qui a du être amplifiée par le système de surexpression que nous avons utilisé. En résumé, nous suggérons que Aah est une protéine qui atteint un équilibre entre le cytoplasme et la périphérie de la membrane interne.

Comme mentionné précédemment, Aah et ses homologues, dont TibC, présentent une certaine similarité de séquence avec les heptosyltransférases impliquées dans la biosynthèse des LPS. Comme la structure de deux heptosyltransférases de *E. coli*, soit WaaC et RfaF, sont connues, nous avons généré un modèle structural de Aah basé sur un alignement avec ces deux protéines (résultats supplémentaires). La structure prédite par le modèle est une structure classique des glycosyltransférases possédant un repliement de type GT-B, c'est-à-dire possédant deux domaines avec un repliement de type Rossmann séparés par une région de jonction (363). Le domaine C-terminal de Aah est celui qui montre la plus forte similarité avec WaaC, ce qui n'est pas surprenant puisqu'il correspond au domaine de liaison avec le résidu heptose (362). Basé sur l'alignement de ce modèle avec la structure connue de WaaC, nous avons identifié trois résidus essentiels pour la fonction de WaaC et conservés dans Aah, soit Asp25, Lys228 et Glu254. La mutation d'un de ces acides aminés entraîne une abolition complète de la fonctionnalité de Aah (résultats supplémentaires). Toutefois, un motif essentiel pour la fonction de WaaC nommé «signature des heptosyltransférases» [D(S/T)(G/A)xxH] et strictement conservé chez toutes les

heptosyltransférases connues (362), incluant l'enzyme OpsX de *H. influenzae* qui présente une spécificité de substrat différente (365), n'est pas retrouvé dans la séquence de Aah. De plus, le domaine N-terminal de Aah possède très peu de similarité avec WaaC ou RfaF et ne peut être modélisé à partir de ces deux structures. Aah et ses homologues ciblent des protéines au lieu d'une molécule lipidique telle que le Kdo<sub>2</sub>-LipideA, ce qui peut expliquer que certains éléments essentiels des heptosyltransférases associées aux LPS ne sont pas retrouvés chez Aah. Nous suggérons donc que Aah est le prototype d'une nouvelle famille d'heptosyltransférases, qui conservent en partie uniquement le mécanisme de reconnaissance et de transfert des sucres commun avec les heptosyltransférases précédemment décrites.

Pour la plupart des systèmes de O-glycosylation, la question de la reconnaissance du substrat est le sujet de nombreuses études. Malgré tout, à ce jour, aucun motif consensus général n'a pu être identifié. Pour certains systèmes de O-glycosylation, il est suggéré que l'oligosaccharyltransférase reconnaît un élément de structure. C'est le cas pour la flagelline de *C. jejuni* où les sucres sont retrouvés au niveau de surfaces exposées dans la structure native (211). Dans d'autres systèmes, la glycosylation a lieu dans des régions de faible complexité (235) ou enrichie en certains types d'acides aminés tels que la thréonine et la proline (242), suggérant que les acides aminés entourant le site de glycosylation jouent aussi un rôle important dans la reconnaissance par la glycosyltransférase. Pour AIDA-I, nous avons démontré que la glycosylation se fait au niveau des séquences répétées de 19 acides aminés retrouvées dans le domaine passager, séquences qui forment la base de l'hélice  $\beta$  prédite en adoptant une structure particulière de trois petits brins  $\beta$  (article 3). Malgré le fait que tous les sucres sont localisés dans cette région répétée de la protéine, aucun motif requis pour la glycosylation n'a pu être identifié par alignement des séquences primaires.

Afin de déterminer le mécanisme de reconnaissance du substrat par Aah, nous avons essayé de trouver le plus petit peptide provenant d'AIDA-I requis pour permettre la glycosylation d'une protéine hétérologue. Ce dernier correspond à un seul motif de 19 acides aminés et peut même être réduit à 13 acides aminés sans affecter le niveau de glycosylation (article 6). Ce motif nous a permis d'entreprendre une étude de mutagenèse pour constater que la reconnaissance ne se fait pas par la séquence primaire en acides aminés, mais plutôt via un motif structural. Ce motif, nommé cassette de glycosylation, correspond à brin  $\beta$  - boucle -

brin  $\beta$ , où la sérine modifiée doit être au début de la boucle. La reconnaissance d'un motif structural a aussi été suggérée pour la  $\beta$ -*N*-acetylglucosamine transférase, une O-glycosyltransférase humaine (366). Cette observation est importante puisque pour la plupart des O-glycosyltrasférases bactériennes, aucune séquence primaire ne semble suffire pour expliquer le mécanisme de reconnaissance. La reconnaissance d'un motif structural fournit aussi une explication pour l'hétérogénéité de la glycosylation et la grande variété de sites modifiés (article 3).

Tel que démontré, la glycosylation de AIDA-I a lieu dans le cytoplasme, avant que la protéine ne traverse les deux membranes constituant l'enveloppe bactérienne. Les substrats pour le système Sec sont transportés sous une forme dépliée et le mécanisme de sécrétion à travers la membrane externe des autotransporteurs monomériques ne permet qu'un niveau minimal de repliement. Toutefois, le motif de reconnaissance de Aah est un élément structural, donc un certain niveau de repliement doit avoir lieu dans le cytoplasme. Comment rassembler ces deux éléments qui semblent s'opposer ?

Le motif que nous suggérons comme site de reconnaissance correspond à une structure secondaire simple (brins  $\beta$  de 4 ou 5 acides aminés). De plus, comme une séquence de seulement 19 acides aminés est suffisante pour permettre la glycosylation, cet élément de structure est probablement minimal. En fait, on suggère que Aah reconnaît un élément structural local et non pas la structure finale bien ordonnée d'hélice  $\beta$  prédite pour AIDA-I. En effet, il est fort probable que la structure tertiaire finale ne soit pas suffisamment flexible pour permettre la fixation de Aah et l'ajout des sucres. En accord, nous n'avons vu aucune glycosylation de TibA non-glycosylé natif en présence de TibC dans un test de glycosylation *in vitro* (résultats supplémentaires).

Un autre élément de réponse à cette question provient des expériences de glycosylation *in vitro* avec Aah et TibC. Nous avons remarqué que la glycosylation *in vitro* d'AIDA-I est dépendante de la présence d'un système de transcription/traduction, suggérant ainsi une glycosylation co-traductionnelle. Cette observation est possiblement la clé du mécanisme d'action de cette famille de glycosyltransférases bactériennes. Les récents développements dans l'étude du mécanisme de traduction ont changé notre façon de voir le ribosome. Il est maintenant clair que le ribosome joue aussi un rôle actif dans l'initiation du repliement de plusieurs protéines. En effet, les polypeptides nouvellement synthétisés peuvent commencer à acquérir des éléments de structure simples avant même que ces derniers

émergent du ribosome (367-370). Nous avons aussi observé une interaction entre la protéine cytoplasmique AIDA\*, Aah et la protéine ribosomale S2 de la sous-unité 30S du ribosome (résultats supplémentaires). Dans cette optique, il est tentant de suggérer que Aah reconnaît une structure particulière qui se forme à la sortie du ribosome, expliquant la dépendance entre la traduction et la glycosylation. En plus, une glycosylation co-traductionnelle fournit un mécanisme plutôt élégant permettant de coordonner la traduction et la glycosylation avec la sécrétion à travers la membrane interne.

Les chaînes naissantes qui sortent du ribosome se retrouvent dans l'environnement particulièrement encombré qu'est le cytoplasme et, dans ces conditions, la tendance de ces protéines à se replier dans un état non-natif et/ou à agréger est plus élevée, pouvant causer des effets néfastes pour les bactéries (371). Cette tendance est particulièrement importante pour les protéines riches en brins  $\beta$  telles que les fibres amyloïdes, des protéines possédant des répétitions ordonnées et adoptant une structure en forme de fibre riche en brins  $\beta$  (372). Une protéine appartenant à la famille des SAATs, soit l'Ag43, a été décrite comme formant des structures de type amyloïde (233) et AIDA-I possède une structure particulièrement enrichie en brins  $\beta$ . Dans cette optique, le rôle de la glycosylation co-traductionnelle pourrait être double. En effet, en plus de permettre la glycosylation complète avant le transport par le système Sec, l'ajout des sucres dès que le polypeptide sort du ribosome pourrait limiter le risque d'agrégation non-fonctionnelle. Il a d'ailleurs été suggéré que l'ajout de sucres sur une protéine permet une meilleure stabilité, non pas en stabilisant un état natif, mais plutôt en déstabilisant un état non-replié (373). En fait, la présence des sucres peut possiblement forcer les polypeptides non-repliés à adopter une conformation plus étendue, évitant ainsi une agrégation non-productive. Dans le cas de AIDA-I, qui possède une structure riche en brins  $\beta$  et la capacité d'interagir avec elle-même avec une haute affinité, il peut être tentant de suggérer que la présence des sucres pourrait aider au maintien de AIDA-I dans un état compatible avec la sécrétion.

## **2) Recrutement d'AIDA-I au complexe SecYEG**

Beaucoup d'études ont déjà été publiées au sujet du recrutement des autotransporteurs monomériques possédant l'ESPR au complexe SecYEG. Toutefois, la plupart de ses études ont été faites avec des autotransporteurs appartenant à la famille des SPATEs, et malgré tout, un modèle consensus n'est toujours pas établi. AIDA-I est aussi retrouvé chez *E. coli*,

mais possède une biogenèse légèrement différente, c'est-à-dire qu'il doit y avoir une coordination entre la traduction, glycosylation et sécrétion. Dans cette optique, nous avons voulu voir si l'extension N-terminale de la séquence signal d'AIDA-I jouait un rôle important dans le processus de sécrétion et de glycosylation. Comme démontré dans l'article 4, la délétion de l'ESPR cause une diminution du niveau d'expression, ce qui est contraire aux études faites avec Pet (292) et Hbp (290). La protéine retrouvée à la surface est glycosylée, suggérant que *a priori*, l'ESPR n'est pas absolument essentielle pour la glycosylation. Toutefois, ces résultats peuvent aussi être interprétés en considérant le rôle de la glycosylation pour AIDA-I. En effet, la forme non-glycosylée d'AIDA-I n'est pas stable et difficilement détectable. Il se pourrait donc que l'absence de l'ESPR entraîne une diminution de l'efficacité de glycosylation. Par conséquent, une partie des polypeptides nouvellement sécrétés seraient retrouvés sous forme non-glycosylée et donc sujet à la dégradation. En suivant ce raisonnement, nos résultats ne sont pas différents de ceux décrits précédemment, c'est-à-dire que la délétion de l'ESPR n'a pas d'effet sur le niveau de protéine transporté par le système Sec, mais plutôt sur l'efficacité de glycosylation. Dans le même ordre d'idée, il a été suggéré que la présence de l'ESPR permettrait un ralentissement de la sécrétion à travers la membrane interne (288, 291). Dans le cas d'AIDA-I, un tel ralentissement pourrait être bénéfique pour permettre une glycosylation complète des polypeptides dans le cytoplasme avant la sécrétion. D'ailleurs, tous les autotransporteurs ou TPS décrits comme étant glycosylés possèdent cette extension N-terminale de la séquence signal.

L'implication de facteurs cytoplasmiques dans le recrutement des autotransporteurs monomériques est aussi un sujet de débat. Comme mentionné dans l'introduction, certains autotransporteurs comme Hbp semblent utiliser de façon facultative la voie SecB ou SRP-dépendante (290). D'un autre côté, il a été suggéré pour Pet et EspP que l'ESPR de ces séquences signal pourrait permettre le recrutement d'un facteur cytoplasmique non-identifié (288, 289). Dans les travaux présentés ici, nous n'avons pas abordé ce problème directement, mais nous avons toutefois fait une observation intéressante. En effet, En purifiant la protéine AIDA\* bloquée dans le cytoplasme, nous avons co-purifié DnaK (résultats supplémentaires). DnaK est considérée comme une chaperonne cytoplasmique générale. Ce résultat est intéressant puisque DnaK a déjà été décrit comme interagissant spécifiquement avec les protéines destinées à la membrane externe pour prolonger leur maintien dans le cytoplasme dans une conformation compatible avec leur transport par le

système Sec (61). De plus, DnaK a été impliquée dans la biogenèse des curli, en prévenant l'agrégation prématurée de CsgA et la formation de fibres amyloïdes dans le cytoplasme (374). Finalement, DnaK est aussi requise pour la localisation polaire cytoplasmique de IscA et SepA, deux autotransporteurs monomériques (334). Quoique nous ne possédons aucune preuve expérimentale directe liant AIDA-I avec DnaK, l'idée que cette chaperonne joue un rôle dans la biogenèse d'au moins quelques autotransporteurs monomériques mérite d'être considérée.

### **3) Clivage protéolytique d'AIDA-I**

Beaucoup d'autotransporteurs monomériques sont clivés suite à leur sécrétion à travers la membrane externe permettant ainsi leur relâche dans le milieu extracellulaire. Différents mécanismes de clivage, autocatalytique ou non, ont déjà été décrits dans la littérature (375). Le domaine passager d'AIDA-I est clivé mais reste associé à la membrane externe par une liaison non-covalente. En mettant au point une technique de purification d'AIDA-I natif, nous avons observé que le domaine passager AIDA-I mature et AIDAc (domaine  $\beta$  et région de jonction) sont co-purifiés, ce qui suggère que malgré le clivage, ces deux domaines de la protéine restent étroitement associés (article 1). Cette observation a été confirmée par nos résultats démontrant que la seule façon de séparer ces deux fragments est par la dénaturation complète du domaine passager.

Nous avons aussi identifié le mécanisme de clivage d'AIDA-I (article 5). Ce mécanisme, qui n'avait pas été décrit auparavant dans la famille des autotransporteurs monomériques, implique deux résidus avec un groupement carboxyle présents dans la région de jonction. Ce mécanisme de clivage pour un autotransporteur est une nouveauté et la découverte d'un tel mécanisme peut aussi contribuer à notre compréhension de la biogenèse d'AIDA-I. En regardant le modèle structural englobant le site de clivage d'AIDA-I, on peut facilement remarquer que les deux résidus impliqués dans le clivage doivent être parfaitement alignés avec leur site cible, ce qui concorde avec notre observation que le repliement et le clivage d'AIDA-I sont deux événements étroitement liés. Toutefois, on retrouve la protéine précurseur à la surface de la bactérie (AIDA-I avant le clivage). Cette observation suggère donc que la protéine, incluant la région de jonction, atteint sa conformation finale uniquement lorsque le polypeptide entier est rendu dans l'espace extracellulaire. Ce mécanisme de clivage, qui nécessite un état de repliement avancé, si ce n'est pas final,



permet aussi d'expliquer la forte association qui existe entre AIDA-I mature et AIDA<sub>c</sub>. En effet, au moment du clivage, la protéine est probablement déjà repliée sous la forme d'une hélice  $\beta$ , soit une structure tertiaire stable qui implique la formation de nombreuses liaisons hydrogènes entre les différents brins  $\beta$ . Dans ce contexte, on peut facilement comprendre pourquoi la dissociation d'AIDA-I mature et d'AIDA<sub>c</sub> requiert la dénaturation complète du domaine passager.

Malgré l'élégance de ce modèle, la question du rôle du clivage demeure. En effet, dans nos études, autant en modifiant le site de clivage (article 1) ou les acides aminés catalytiques (article 5), nous n'avons remarqué aucune différence majeure du point de vue de la fonction ou de la stabilité entre AIDA-I natif et AIDA-I non-clivé. Toutefois, étant donné la diversité des mécanismes de clivage existants ainsi que le nombre élevé d'autotransporteurs monomériques clivés, il serait quand même assez surprenant que le clivage d'AIDA-I ne soit qu'un vestige de l'évolution et n'ait aucune fonction réelle. Le système que nous utilisons est un système de surexpression. Il est donc possible que dans ces conditions, l'effet de l'abolition du clivage soit compensé par la surexpression. De plus, nous avons testé les différents phénotypes connus associés à l'expression d'AIDA-I *in vitro* uniquement. Il ne faut toutefois pas exclure un possible rôle du clivage dans le cas d'une infection *in vivo*. AIDA-I est une protéine hautement glycosylée par des sucres uniques aux bactéries et les sucres sont reconnus pour être des composants qui peuvent activer le système immunitaire. Dans ce contexte, la relâche du domaine passager glycosylé pourrait être une façon de contourner la réponse immunitaire. Parallèlement, il a été démontré pour l'autotransporteur Hap de *H. influenzae*, impliqué principalement dans l'adhésion et l'auto-agrégation bactérienne, que la relâche du domaine passager permettait un contrôle de ces deux fonctions (331). Donc, la relâche d'une adhésine dans le milieu extracellulaire peut servir une fonction particulière. De plus, d'un point de vue conceptuel, la relâche du domaine passager pourrait aussi être requise pour permettre à une bactérie individuelle de se séparer d'un agrégat bactérien ou d'un biofilm afin de se disperser et éventuellement coloniser un nouvel environnement.

#### 4) Fonctionnalité

Comme mentionné précédemment, 97% des autotransporteurs monomériques sont prédits pour posséder une hélice  $\beta$  comme structure finale (264). Pourtant, une variété de fonctions a été attribuée aux domaines passagers d'autotransporteurs. De plus, plusieurs de ces protéines sont multifonctionnelles, ce qui signifie que diverses fonctions peuvent être associées à leur expression (269). Dans le cas d'AIDA-I, cette hélice  $\beta$  serait principalement constituée des 35 répétitions du motif imparfait de 19 acides aminés retrouvés dans le domaine passager. Aucun domaine ou boucle sortant de cette hélice  $\beta$  ne peut être prédit à partir de la séquence primaire. Toutefois, pour plusieurs autotransporteurs, l'hélice  $\beta$  semble servir uniquement d'échafaud pour l'exposition de domaines fonctionnels (4). Malgré le fait que la structure prédite d'AIDA-I semble être une hélice  $\beta$  plutôt uniforme, cette protéine est impliquée dans diverses fonctions, telles que l'adhésion et l'invasion de cellules épithéliales ainsi que la formation de biofilm et d'agrégats bactériens (231, 338). Pour cette raison, nous avons décidé d'entreprendre une étude de structure-fonction du domaine passager d'AIDA-I par une approche de mutagenèse aléatoire afin d'identifier des régions importantes pour la fonctionnalité de cette protéine (article 2).

À partir de ces résultats, notre principale observation est que la région N-terminale d'AIDA-I mature (environ les 225 premiers acides aminés après la séquence signal) est importante pour le phénotype d'adhésion uniquement. Ce résultat est logique puisque la plupart des adhésines bactériennes sont bâties de façon à exposer le domaine adhésion le plus loin possible de la surface bactérienne afin d'optimiser les contacts avec un récepteur cellulaire. Nous avons aussi identifié un deuxième domaine possiblement impliqué dans l'adhésion, un peu plus en C-terminal d'AIDA-I mature. Un domaine d'adhésion situé plus près de la membrane pourrait éventuellement permettre une adhésion plus intime entre la bactérie et les cellules. Deux domaines d'adhésion sur la même protéine n'est pas unique à AIDA-I et peut être avantageux pour une bactérie. En effet, dans le cas de l'autotransporteur Hia de *H. influenzae*, les deux domaines permettent l'adhésion avec le même récepteur, mais avec une affinité différente (376). Dans le cas d'AIDA-I, les deux domaines se situent sur deux faces de l'hélice  $\beta$  prédite. Cette disposition pourrait permettre à AIDA-I d'effectuer des interactions multivalentes avec un récepteur potentiel. Alternativement, AIDA-I pourrait se fixer à deux récepteurs différents. Toutefois, comme

aucun récepteur pour AIDA-I n'a encore été identifié, l'interprétation de ces résultats reste au stade d'hypothèses uniquement.

Avec cette approche de mutagenèse, nous avons pu séparer les fonctions associées à l'expression d'AIDA-I. En effet, un variant incapable de former des agrégats bactériens peut toujours permettre la formation de biofilm et vice versa. Des résultats similaires ont été obtenus pour l'adhésion et l'invasion. Parallèlement, des études similaires ont été effectuées avec les deux autres membres de la famille des SAATs, soit l'Ag43 et TibA. L'étude effectuée avec l'Ag43 a permis l'identification de régions impliquées dans l'agrégation et la formation de biofilm (377), tandis que celle faite avec TibA a permis la séparation de régions impliquées dans l'adhésion et l'agrégation (378).

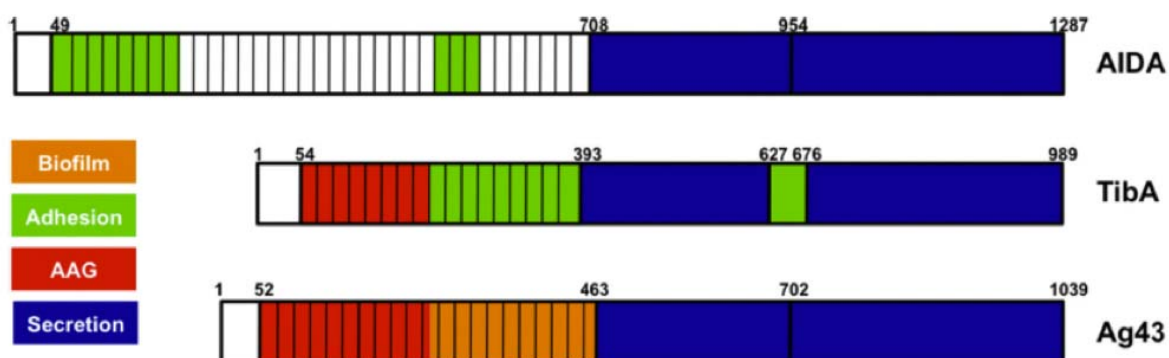


Figure 1 : Organisation modulaire des SAATs [figure adaptée de la référence (378)]. Représentation schématique des trois membres de la famille des SAATs, soit AIDA-I, l'Ag43 et TibA. Les régions connues comme étant responsable des fonctions suivantes sont indiquées : bleu, sécrétion; orange, biofilm; vert, adhésion; rouge, AAG ou agrégation. Aucun domaine précis n'a été identifié pour le phénotype d'invasion.

En couplant ces trois études de structure-fonction sur les SAATs, il apparaît de plus en plus que ces protéines sont organisées sous forme modulaire. Cette observation est importante puisqu'il a longtemps été supposé que ces protéines adoptaient une conformation en hélice  $\beta$  plutôt rigide et uniforme. De plus, le fait que plusieurs fonctions différentes peuvent être associées à une région de ces protéines qui est, du point de vue de la séquence primaire, très similaire, est aussi une observation importante pour notre compréhension du mode de fonctionnement de ces autotransporteurs. En effet, dans la situation présente, on ne peut pas définir le rôle d'une protéine en se basant uniquement sur la présence de ce motif répété de 19 acides aminés (donc, sur la séquence primaire du domaine passager). Pour l'instant,

AIDA-I, l'Ag43 et TibA sont les seuls membres décrits appartenant à la famille des SAATs et partageant le motif répété de 19 acides aminés. Toutefois, en recherchant de homologues de Aah, nous avons observé que plusieurs glycosyltransférases putatives sont associées à une protéine qui possède toutes les caractéristiques d'un autotransporteur monomérique en plus du motif imparfait de 19 acides aminés. La taille de ces autotransporteurs putatifs varie énormément et le nombre de répétition du motif conservé peut aller jusqu'à 93 dans le cas de l'autotransporteur retrouvé chez *C. sakasakii* (article 6). Comme ces autotransporteurs putatifs ont tous une longueur différente, il est plutôt tentant de supposer que toutes ces protéines proviennent d'un assemblage de « modules » fonctionnels différents. L'acquisition ou la perte de « modules » fonctionnels pourrait être simplement reliée aux besoins ou pressions exercées sur les bactéries en fonction de l'environnement dans lequel elles se retrouvent.

## Conclusion

Les résultats présentés dans cette thèse apportent de nouvelles informations et permettent une meilleure compréhension de la biogenèse d'une des plus importantes familles de protéines sécrétées chez les bactéries Gram négatif, soit les autotransporteurs monomériques.

À ce titre, nous avons mis en évidence la subdivision du domaine passager d'AIDA-I en « modules » fonctionnels, remettant ainsi en question notre façon de voir les domaines passagers d'autotransporteurs. En effet, nos résultats indiquent que les autotransporteurs ont évolué de façon à maximiser leur organisation, c'est-à-dire que leurs domaines fonctionnels peuvent directement avoir la structure d'hélice  $\beta$  requise pour la sécrétion à travers la membrane externe. Afin d'appuyer cette hypothèse, la caractérisation d'autres autotransporteurs monomériques, principalement ceux retrouvés chez des espèces bactériennes phylogénétiquement éloignées de *E. coli*, serait requise. Toutefois, l'importance de chacun de ces « modules » fonctionnels dans un contexte d'infection n'a jamais été réellement étudiée. Est-ce que l'ajout de « modules » ne sert qu'à exposer le domaine d'adhésion le plus loin possible de la surface bactérienne ou est-ce que la formation de biofilm ou d'agrégats bactériens procure un avantage significatif aux bactéries pour l'établissement d'une infection ou pour la dissémination? Pour poursuivre dans cette voie, les variants générés dans cette étude et qui sont déficients pour une seule des fonctions associées à l'expression d'AIDA-I pourraient être utilisés. Toutefois, à ce stade, l'implantation d'un modèle d'infection animal est primordiale.

Un autre aspect de la biogenèse d'AIDA-I sur lequel nous avons apporté de nouvelles informations est la glycosylation. Nous avons démontré que la glycosylation est essentielle pour la stabilité d'AIDA-I à la surface bactérienne et, par conséquent, pour la fonction d'adhésion reliée à cette protéine. Par ailleurs, la glycosylation ne semble pas être requise pour la capacité des bactéries exprimant AIDA-I à former des biofilms ou des agrégats bactériens. Comme seul le phénotype d'adhésion est influencé par une diminution du niveau d'expression de la protéine, est-ce que la glycosylation ne pourrait pas servir de mécanisme de contrôle pour coordonner les différentes fonctions reliées à l'expression d'AIDA-I? Dans cette optique, l'arrêt de la glycosylation pourrait servir de « signal » de transition entre la fonction d'adhésion aux cellules et la formation agrégats bactériens. Une

étude plus ciblée du mécanisme de régulation de l'opéron *aah-aidA* serait bénéfique pour vérifier cette hypothèse.

Par l'étude de la glycosylation d'AIDA-I, nous avons aussi mis en évidence une nouvelle famille de glycosyltransférases bactériennes, représentée par Aah. Les membres de cette famille sont présents chez une variété de protéobactéries et plusieurs d'entre eux sont retrouvés en association avec des protéines qui possèdent tous les aspects caractéristiques d'autotransporteurs monomériques ou trimériques. Cette association suggère que ces autotransporteurs pourraient aussi être modifiés par l'ajout de sucres. Tout comme nous l'avons démontré pour AIDA-I, cette glycosylation pourrait avoir des conséquences insoupçonnées sur la fonctionnalité ou la biogenèse de ces autotransporteurs. Il serait donc important d'étudier l'effet de la glycosylation sur ces autres autotransporteurs afin de mieux comprendre l'impact de la glycosylation sur les protéines exposées à la surface des bactéries.

En caractérisant cette nouvelle famille de glycosyltransférases, nous avons décrit le premier système de O-glycosylation cytoplasmique chez les bactéries Gram négatif. De plus, nous avons démontré que la glycosylation est une modification co-translationnelle, ce qui est sans précédent chez les bactéries. Une glycosylation co-translationnelle représente un moyen efficace pour coordonner la traduction, la glycosylation ainsi que le transport à travers la membrane interne. Toutefois, selon moi, cette observation génère plus de questions que de réponses. Tout d'abord, est-ce que Aah peut interagir directement avec le ribosome? Si oui, comment est coordonnée l'interaction entre les différentes chaperonnes cytoplasmiques (TF, SRP, DnaK, DnaJ et autres) présentes à la sortie du ribosome, Aah et la chaîne polypeptidique naissante? Plusieurs structures de haute résolution du complexe ribosomal sont maintenant disponibles, permettant d'obtenir une image en trois dimensions du ribosome lors du processus de traduction (379, 380). Dans le cas où Aah entrerait directement en interaction avec le ribosome, l'identification de la ou des protéines ribosomales responsables de cette interaction permettrait de mieux comprendre l'ordre des événements à la sortie du ribosome. Toutefois, une glycosylation co-translationnelle n'implique pas nécessairement une interaction directe avec le ribosome. Dans ce cas, est-ce qu'un élément de séquence permettrait le recrutement de Aah au niveau du polypeptide en cours de traduction? L'élément le plus logique serait l'extension N-terminale de la séquence signal. En accord avec cette hypothèse, la délétion de l'ESPR d'AIDA-I a un effet marqué sur le niveau de protéine à la surface. Similairement, la délétion de l'ESPR de

l'autotransporteur trimérique glycosylé EmaA de *Aggregatibacter actinomycetemcomitans* affecte dramatiquement la sécrétion de cette protéine (381). Dans cette optique, une étude plus approfondie du rôle de cette extension de la séquence signal pour les glycoprotéines sécrétées par le système de sécrétion de type V est requise.

Les perspectives et les répercussions de ce projet sont nombreuses. En effet, peu de modifications protéiques co-traductionnelles sont connues chez les bactéries et notre connaissance de leurs mécanismes d'action est très limitée (380). De plus, les modifications connues, telles que l'enlèvement du groupe formyle de la méthionine N-terminale ainsi que l'excision de cette dernière, ciblent uniquement l'extrémité N-terminale du polypeptide. Dans cette optique, la glycosylation effectuée par Aah représente un nouveau type de modification co-traductionnelle chez les bactéries et une meilleure connaissance du mécanisme d'action de cette enzyme pourrait fournir de précieux renseignements, non seulement sur le mode d'action de cette famille de glycosyltransférases mais aussi sur le processus de traduction chez les bactéries.

Le prototype de cette famille de glycosyltransférases et le membre le mieux caractérisé, Aah, est retrouvé chez *E. coli*. La découverte d'une telle enzyme permettant l'ajout des sucres sur une protéine directement chez *E. coli* est une avancée importante d'un point de vue biotechnologique. En effet, *E. coli* est une bactérie facilement cultivable et à faible coût d'utilisation. De plus, comme Aah agit directement dans le cytoplasme, tous les problèmes liés au transport dans le périplasma d'éventuelles protéines recombinantes sont éliminés. Il ne reste qu'à savoir si la spécificité de Aah pour le sucre utilisé peut être modifiée de façon à produire des protéines d'origine eucaryote glycosylées avec leurs sucres natifs dans *E. coli*.

**Bibliographie**

1. Waksman G. 2012. Bacterial secretion comes of age. *Philos Trans R Soc Lond B Biol Sci* 367: 1014-5
2. Economou A, Christie PJ, Fernandez RC, Palmer T, Plano GV, Pugsley AP. 2006. Secretion by numbers: Protein traffic in prokaryotes. *Mol Microbiol* 62: 308-19
3. Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D. 2004. Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* 68: 692-744
4. Leyton DL, Rossiter AE, Henderson IR. 2012. From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nat Rev Microbiol* 10: 213-25
5. Klemm P, Vejborg RM, Sherlock O. 2006. Self-associating autotransporters, SAATs: functional and structural similarities. *Int J Med Microbiol* 296: 187-95
6. Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2: 123-40
7. Benz I, Schmidt MA. 2001. Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Mol Microbiol* 40: 1403-13
8. Lindenthal C, Elsinghorst EA. 1999. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect Immun* 67: 4084-91
9. Sherlock O, Dobrindt U, Jensen JB, Munk Vejborg R, Klemm P. 2006. Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. *J Bacteriol* 188: 1798-807
10. Szymanski CM, Wren BW. 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* 3: 225-37
11. Finlay BB, Falkow S. 1997. Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev* 61: 136-69
12. Salmond GP, Reeves PJ. 1993. Membrane traffic wardens and protein secretion in gram-negative bacteria. *Trends Biochem Sci* 18: 7-12
13. Henderson IR, Nataro JP, Kaper JB, Meyer TF, Farrand SK, Burns DL, Finlay BB, St Geme JW, 3rd. 2000. Renaming protein secretion in the gram-negative bacteria. *Trends Microbiol* 8: 352



14. du Plessis DJ, Nouwen N, Driessen AJ. 2011. The Sec translocase. *Biochim Biophys Acta* 1808: 851-65
15. Pohlschroder M, Prinz WA, Hartmann E, Beckwith J. 1997. Protein translocation in the three domains of life: variations on a theme. *Cell* 91: 563-6
16. Papanikou E, Karamanou S, Economou A. 2007. Bacterial protein secretion through the translocase nanomachine. *Nat Rev Microbiol* 5: 839-51
17. Brundage L, Hendrick JP, Schiebel E, Driessen AJ, Wickner W. 1990. The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* 62: 649-57
18. Cannon KS, Or E, Clemons WM, Jr., Shibata Y, Rapoport TA. 2005. Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *J Cell Biol* 169: 219-25
19. Matsuyama S, Akimaru J, Mizushima S. 1990. SecE-dependent overproduction of SecY in *Escherichia coli*. Evidence for interaction between two components of the secretory machinery. *FEBS Lett* 269: 96-100
20. Homma T, Yoshihisa T, Ito K. 1997. Subunit interactions in the *Escherichia coli* protein translocase: SecE and SecG associate independently with SecY. *FEBS Lett* 408: 11-5
21. Nishiyama K, Hanada M, Tokuda H. 1994. Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of *Escherichia coli* at low temperature. *Embo J* 13: 3272-7
22. Nishiyama K, Suzuki T, Tokuda H. 1996. Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation. *Cell* 85: 71-81
23. van der Sluis EO, van der Vries E, Berrelkamp G, Nouwen N, Driessen AJ. 2006. Topologically fixed SecG is fully functional. *J Bacteriol* 188: 1188-90
24. Van den Berg B, Clemons WM, Jr., Collinson I, Modis Y, Hartmann E, Harrison SC, Rapoport TA. 2004. X-ray structure of a protein-conducting channel. *Nature* 427: 36-44
25. Gumbart J, Schulten K. 2006. Molecular dynamics studies of the archaeal translocon. *Biophys J* 90: 2356-67
26. Tian P, Andricioaei I. 2006. Size, motion, and function of the SecY translocon revealed by molecular dynamics simulations with virtual probes. *Biophys J* 90: 2718-30

27. Bonardi F, Halza E, Walko M, Du Plessis F, Nouwen N, Feringa BL, Driessen AJ. 2011. Probing the SecYEG translocation pore size with preproteins conjugated with sizable rigid spherical molecules. *Proc Natl Acad Sci U S A* 108: 7775-80
28. Becker T, Bhushan S, Jarasch A, Armache JP, Funes S, Jossinet F, Gumbart J, Mielke T, Berninghausen O, Schulten K, Westhof E, Gilmore R, Mandon EC, Beckmann R. 2009. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* 326: 1369-73
29. Scheuring J, Braun N, Nothdurft L, Stumpf M, Veenendaal AK, Kol S, van der Does C, Driessen AJ, Weinkauff S. 2005. The oligomeric distribution of SecYEG is altered by SecA and translocation ligands. *J Mol Biol* 354: 258-71
30. Hanein D, Matlack KE, Jungnickel B, Plath K, Kalies KU, Miller KR, Rapoport TA, Akey CW. 1996. Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* 87: 721-32
31. Mitra K, Frank J, Driessen A. 2006. Co- and post-translational translocation through the protein-conducting channel: analogous mechanisms at work? *Nat Struct Mol Biol* 13: 957-64
32. Deville K, Gold VA, Robson A, Whitehouse S, Sessions RB, Baldwin SA, Radford SE, Collinson I. 2011. The oligomeric state and arrangement of the active bacterial translocon. *J Biol Chem* 286: 4659-69
33. Mitra K, Schaffitzel C, Shaikh T, Tama F, Jenni S, Brooks CL, 3rd, Ban N, Frank J. 2005. Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* 438: 318-24
34. Sardis MF, Economou A. 2010. SecA: a tale of two protomers. *Mol Microbiol* 76: 1070-81
35. Gouridis G, Karamanou S, Gelis I, Kalodimos CG, Economou A. 2009. Signal peptides are allosteric activators of the protein translocase. *Nature* 462: 363-7
36. Papanikolaou Y, Papadovasilaki M, Ravelli RB, McCarthy AA, Cusack S, Economou A, Petratos K. 2007. Structure of dimeric SecA, the *Escherichia coli* preprotein translocase motor. *J Mol Biol* 366: 1545-57
37. Sharma V, Arockiasamy A, Ronning DR, Savva CG, Holzenburg A, Braunstein M, Jacobs WR, Jr., Sacchettini JC. 2003. Crystal structure of *Mycobacterium tuberculosis* SecA, a preprotein translocating ATPase. *Proc Natl Acad Sci U S A* 100: 2243-8

38. Zimmer J, Li W, Rapoport TA. 2006. A novel dimer interface and conformational changes revealed by an X-ray structure of *B. subtilis* SecA. *J Mol Biol* 364: 259-65
39. Vassylyev DG, Mori H, Vassylyeva MN, Tsukazaki T, Kimura Y, Tahirov TH, Ito K. 2006. Crystal structure of the translocation ATPase SecA from *Thermus thermophilus* reveals a parallel, head-to-head dimer. *J Mol Biol* 364: 248-58
40. Woodbury RL, Hardy SJ, Randall LL. 2002. Complex behavior in solution of homodimeric SecA. *Protein Sci* 11: 875-82
41. de Keyzer J, van der Sluis EO, Spelbrink RE, Nijstad N, de Kruijff B, Nouwen N, van der Does C, Driessen AJ. 2005. Covalently dimerized SecA is functional in protein translocation. *J Biol Chem* 280: 35255-60
42. Or E, Boyd D, Gon S, Beckwith J, Rapoport T. 2005. The bacterial ATPase SecA functions as a monomer in protein translocation. *J Biol Chem* 280: 9097-105
43. Zimmer J, Nam Y, Rapoport TA. 2008. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455: 936-43
44. Sianidis G, Karamanou S, Vrontou E, Boulias K, Repanas K, Kyrpidis N, Politou AS, Economou A. 2001. Cross-talk between catalytic and regulatory elements in a DEAD motor domain is essential for SecA function. *Embo J* 20: 961-70
45. Economou A, Pogliano JA, Beckwith J, Oliver DB, Wickner W. 1995. SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. *Cell* 83: 1171-81
46. van der Wolk JP, de Wit JG, Driessen AJ. 1997. The catalytic cycle of the *Escherichia coli* SecA ATPase comprises two distinct preprotein translocation events. *Embo J* 16: 7297-304
47. Driessen AJ. 1992. Bacterial protein translocation: kinetic and thermodynamic role of ATP and the protonmotive force. *Trends Biochem Sci* 17: 219-23
48. Pogliano JA, Beckwith J. 1994. SecD and SecF facilitate protein export in *Escherichia coli*. *Embo J* 13: 554-61
49. Tsukazaki T, Mori H, Echizen Y, Ishitani R, Fukai S, Tanaka T, Perederina A, Vassylyev DG, Kohno T, Maturana AD, Ito K, Nureki O. 2011. Structure and function of a membrane component SecDF that enhances protein export. *Nature* 474: 235-8

50. Samuelson JC, Chen M, Jiang F, Moller I, Wiedmann M, Kuhn A, Phillips GJ, Dalbey RE. 2000. YidC mediates membrane protein insertion in bacteria. *Nature* 406: 637-41
51. Nouwen N, Driessen AJ. 2002. SecDFyajC forms a heterotetrameric complex with YidC. *Mol Microbiol* 44: 1397-405
52. Kol S, Nouwen N, Driessen AJ. 2008. Mechanisms of YidC-mediated insertion and assembly of multimeric membrane protein complexes. *J Biol Chem* 283: 31269-73
53. Valent QA, Kendall DA, High S, Kusters R, Oudega B, Luirink J. 1995. Early events in preprotein recognition in *E. coli*: interaction of SRP and trigger factor with nascent polypeptides. *Embo J* 14: 5494-505
54. Eisner G, Koch HG, Beck K, Brunner J, Muller M. 2003. Ligand crowding at a nascent signal sequence. *J Cell Biol* 163: 35-44
55. Poritz MA, Bernstein HD, Strub K, Zopf D, Wilhelm H, Walter P. 1990. An *E. coli* ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. *Science* 250: 1111-7
56. Halic M, Beckmann R. 2005. The signal recognition particle and its interactions during protein targeting. *Curr Opin Struct Biol* 15: 116-25
57. Braig D, Mircheva M, Sachelaru I, van der Sluis EO, Sturm L, Beckmann R, Koch HG. 2011. Signal sequence-independent SRP-SR complex formation at the membrane suggests an alternative targeting pathway within the SRP cycle. *Mol Biol Cell* 22: 2309-23
58. Bechtluft P, van Leeuwen RG, Tyreman M, Tomkiewicz D, Nouwen N, Tepper HL, Driessen AJ, Tans SJ. 2007. Direct observation of chaperone-induced changes in a protein folding pathway. *Science* 318: 1458-61
59. Fekkes P, van der Does C, Driessen AJ. 1997. The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *Embo J* 16: 6105-13
60. Ullers RS, Luirink J, Harms N, Schwager F, Georgopoulos C, Genevaux P. 2004. SecB is a bona fide generalized chaperone in *Escherichia coli*. *Proc Natl Acad Sci USA* 101: 7583-8
61. Qi HY, Hyndman JB, Bernstein HD. 2002. DnaK promotes the selective export of outer membrane protein precursors in SecA-deficient *Escherichia coli*. *J Biol Chem* 277: 51077-83

62. Sakr S, Cirinesi AM, Ullers RS, Schwager F, Georgopoulos C, Genevaux P. 2010. Lon protease quality control of presecretory proteins in *Escherichia coli* and its dependence on the SecB and DnaJ (Hsp40) chaperones. *J Biol Chem* 285: 23506-14
63. Paetzel M, Karla A, Strynadka NC, Dalbey RE. 2002. Signal peptidases. *Chem Rev* 102: 4549-80
64. Schatz PJ, Beckwith J. 1990. Genetic analysis of protein export in *Escherichia coli*. *Annu Rev Genet* 24: 215-48
65. Gierasch LM. 1989. Signal sequences. *Biochemistry* 28: 923-30
66. Huber D, Boyd D, Xia Y, Olma MH, Gerstein M, Beckwith J. 2005. Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. *J Bacteriol* 187: 2983-91
67. Peterson JH, Woolhead CA, Bernstein HD. 2003. Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J Biol Chem* 278: 46155-62
68. Ulbrandt ND, Newitt JA, Bernstein HD. 1997. The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* 88: 187-96
69. Patzelt H, Rudiger S, Brehmer D, Kramer G, Vorderwulbecke S, Schaffitzel E, Waitz A, Hesterkamp T, Dong L, Schneider-Mergener J, Bukau B, Deuerling E. 2001. Binding specificity of *Escherichia coli* trigger factor. *Proc Natl Acad Sci U S A* 98: 14244-9
70. Hoffmann A, Bukau B, Kramer G. 2010. Structure and function of the molecular chaperone Trigger Factor. *Biochim Biophys Acta* 1803: 650-61
71. Berndt U, Oellerer S, Zhang Y, Johnson AE, Rospert S. 2009. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proc Natl Acad Sci U S A* 106: 1398-403
72. Bornemann T, Jockel J, Rodnina MV, Wintermeyer W. 2008. Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nat Struct Mol Biol* 15: 494-9
73. Pool MR. 2009. A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. *J Cell Biol* 185: 889-902

74. Berks BC. 1996. A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol* 22: 393-404
75. Hatzixanthis K, Palmer T, Sargent F. 2003. A subset of bacterial inner membrane proteins integrated by the twin-arginine translocase. *Mol Microbiol* 49: 1377-90
76. Rose RW, Bruser T, Kissinger JC, Pohlschroder M. 2002. Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* 45: 943-50
77. Weiner JH, Bilous PT, Shaw GM, Lubitz SP, Frost L, Thomas GH, Cole JA, Turner RJ. 1998. A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell* 93: 93-101
78. Robinson C, Bolhuis A. 2004. Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim Biophys Acta* 1694: 135-47
79. Graubner W, Schierhorn A, Bruser T. 2007. DnaK plays a pivotal role in Tat targeting of CueO and functions beside SlyD as a general Tat signal binding chaperone. *J Biol Chem* 282: 7116-24
80. Genest O, Seduk F, Ilbert M, Mejean V, Iobbi-Nivol C. 2006. Signal peptide protection by specific chaperone. *Biochem Biophys Res Commun* 339: 991-5
81. Jongbloed JD, Grieger U, Antelmann H, Hecker M, Nijland R, Bron S, van Dijl JM. 2004. Two minimal Tat translocases in *Bacillus*. *Mol Microbiol* 54: 1319-25
82. Sargent F, Stanley NR, Berks BC, Palmer T. 1999. Sec-independent protein translocation in *Escherichia coli*. A distinct and pivotal role for the TatB protein. *J Biol Chem* 274: 36073-82
83. McDevitt CA, Buchanan G, Sargent F, Palmer T, Berks BC. 2006. Subunit composition and in vivo substrate-binding characteristics of *Escherichia coli* Tat protein complexes expressed at native levels. *FEBS J* 273: 5656-68
84. Alami M, Luke I, Deitermann S, Eisner G, Koch HG, Brunner J, Muller M. 2003. Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. *Mol Cell* 12: 937-46
85. Gohlke U, Pullan L, McDevitt CA, Porcelli I, de Leeuw E, Palmer T, Saibil HR, Berks BC. 2005. The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc Natl Acad Sci U S A* 102: 10482-6

86. Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J. 2003. Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* 299: 262-5
87. Sklar JG, Wu T, Kahne D, Silhavy TJ. 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes Dev* 21: 2473-84
88. Hagan CL, Silhavy TJ, Kahne D. 2011. beta-Barrel membrane protein assembly by the Bam complex. *Annu Rev Biochem* 80: 189-210
89. Gentle I, Gabriel K, Beech P, Waller R, Lithgow T. 2004. The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. *J Cell Biol* 164: 19-24
90. Paschen SA, Waizenegger T, Stan T, Preuss M, Cyrklaff M, Hell K, Rapaport D, Neupert W. 2003. Evolutionary conservation of biogenesis of beta-barrel membrane proteins. *Nature* 426: 862-6
91. Kim S, Malinverni JC, Sliz P, Silhavy TJ, Harrison SC, Kahne D. 2007. Structure and function of an essential component of the outer membrane protein assembly machine. *Science* 317: 961-4
92. Knowles TJ, Jeeves M, Bobat S, Dancea F, McClelland D, Palmer T, Overduin M, Henderson IR. 2008. Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. *Mol Microbiol* 68: 1216-27
93. Robert V, Volokhina EB, Senf F, Bos MP, Van Gelder P, Tommassen J. 2006. Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol* 4: e377
94. Remaut H, Waksman G. 2006. Protein-protein interaction through beta-strand addition. *Trends Biochem Sci* 31: 436-44
95. Burgess NK, Dao TP, Stanley AM, Fleming KG. 2008. Beta-barrel proteins that reside in the *Escherichia coli* outer membrane in vivo demonstrate varied folding behavior in vitro. *J Biol Chem* 283: 26748-58
96. Kim KH, Aulakh S, Paetzel M. 2011. Crystal structure of beta-barrel assembly machinery BamCD protein complex. *J Biol Chem* 286: 39116-21
97. Knowles TJ, Browning DF, Jeeves M, Maderbocus R, Rajesh S, Sridhar P, Manoli E, Emery D, Sommer U, Spencer A, Leyton DL, Squire D, Chaudhuri RR, Viant MR, Cunningham AF, Henderson IR, Overduin M. 2011. Structure and function of

- BamE within the outer membrane and the beta-barrel assembly machine. *EMBO Rep* 12: 123-8
98. Waksman G, Hultgren SJ. 2009. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol* 7: 765-74
  99. Epstein EA, Chapman MR. 2008. Polymerizing the fibre between bacteria and host cells: the biogenesis of functional amyloid fibres. *Cell Microbiol* 10: 1413-20
  100. Okuda S, Tokuda H. 2011. Lipoprotein sorting in bacteria. *Annu Rev Microbiol* 65: 239-59
  101. Narita S. 2011. ABC transporters involved in the biogenesis of the outer membrane in gram-negative bacteria. *Biosci Biotechnol Biochem* 75: 1044-54
  102. Simeone R, Bottai D, Brosch R. 2009. ESX/type VII secretion systems and their role in host-pathogen interaction. *Curr Opin Microbiol* 12: 4-10
  103. Desvaux M, Parham NJ, Scott-Tucker A, Henderson IR. 2004. The general secretory pathway: a general misnomer? *Trends Microbiol* 12: 306-9
  104. Holland IB, Schmitt L, Young J. 2005. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). *Mol Membr Biol* 22: 29-39
  105. Delepelaire P. 2004. Type I secretion in gram-negative bacteria. *Biochim Biophys Acta* 1694: 149-61
  106. Goebel W, Hedgpeth J. 1982. Cloning and functional characterization of the plasmid-encoded hemolysin determinant of *Escherichia coli*. *J Bacteriol* 151: 1290-8
  107. Ladant D, Ullmann A. 1999. *Bordatella pertussis* adenylate cyclase: a toxin with multiple talents. *Trends Microbiol* 7: 172-6
  108. Kleanthous C. 2010. Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat Rev Microbiol* 8: 843-8
  109. Chervaux C, Holland IB. 1996. Random and directed mutagenesis to elucidate the functional importance of helix II and F-989 in the C-terminal secretion signal of *Escherichia coli* hemolysin. *J Bacteriol* 178: 1232-6
  110. Nishi J, Sheikh J, Mizuguchi K, Luisi B, Burland V, Boutin A, Rose DJ, Blattner FR, Nataro JP. 2003. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. *J Biol Chem* 278: 45680-9



111. Andersen C. 2003. Channel-tunnels: outer membrane components of type I secretion systems and multidrug efflux pumps of Gram-negative bacteria. *Rev Physiol Biochem Pharmacol* 147: 122-65
112. Pugsley AP, Francetic O, Possot OM, Sauvonnnet N, Hardie KR. 1997. Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in Gram-negative bacteria--a review. *Gene* 192: 13-9
113. d'Enfert C, Pugsley AP. 1987. A gene fusion approach to the study of pullulanase export and secretion in *Escherichia coli*. *Mol Microbiol* 1: 159-68
114. Pugsley AP. 1993. The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* 57: 50-108
115. Cianciotto NP. 2009. Many substrates and functions of type II secretion: lessons learned from *Legionella pneumophila*. *Future Microbiol* 4: 797-805
116. Johnson TL, Abendroth J, Hol WG, Sandkvist M. 2006. Type II secretion: from structure to function. *FEMS Microbiol Lett* 255: 175-86
117. Nunn D. 1999. Bacterial type II protein export and pilus biogenesis: more than just homologies? *Trends Cell Biol* 9: 402-8
118. Burrows LL. 2005. Weapons of mass retraction. *Mol Microbiol* 57: 878-88
119. Reichow SL, Korotkov KV, Hol WG, Gonen T. 2010. Structure of the cholera toxin secretion channel in its closed state. *Nat Struct Mol Biol* 17: 1226-32
120. Guilvout I, Chami M, Engel A, Pugsley AP, Bayan N. 2006. Bacterial outer membrane secretin PulD assembles and inserts into the inner membrane in the absence of its pilotin. *Embo J* 25: 5241-9
121. Cornelis GR. 2006. The type III secretion injectisome. *Nat Rev Microbiol* 4: 811-25
122. Galan JE. 2009. Common themes in the design and function of bacterial effectors. *Cell Host Microbe* 5: 571-9
123. Dean P. 2011. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol Rev* 35: 1100-25
124. Worrall LJ, Lameignere E, Strynadka NC. 2011. Structural overview of the bacterial injectisome. *Curr Opin Microbiol* 14: 3-8
125. Blocker A, Jouihri N, Larquet E, Gounon P, Ebel F, Parsot C, Sansonetti P, Allaoui A. 2001. Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton. *Mol Microbiol* 39: 652-63

126. Veenendaal AK, Hodgkinson JL, Schwarzer L, Stabat D, Zenk SF, Blocker AJ. 2007. The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Mol Microbiol* 63: 1719-30
127. Deane JE, Abrusci P, Johnson S, Lea SM. 2010. Timing is everything: the regulation of type III secretion. *Cell Mol Life Sci* 67: 1065-75
128. Gophna U, Ron EZ, Graur D. 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* 312: 151-63
129. Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 3: 711-21
130. Alvarez-Martinez CE, Christie PJ. 2009. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* 73: 775-808
131. Hofreuter D, Odenbreit S, Haas R. 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol Microbiol* 41: 379-91
132. Hamilton HL, Dominguez NM, Schwartz KJ, Hackett KT, Dillard JP. 2005. *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel type IV secretion system. *Mol Microbiol* 55: 1704-21
133. Hubber A, Roy CR. 2010. Modulation of host cell function by *Legionella pneumophila* type IV effectors. *Annu Rev Cell Dev Biol* 26: 261-83
134. Locht C, Coutte L, Mielcarek N. 2011. The ins and outs of pertussis toxin. *FEBS J* 278: 4668-82
135. Rego AT, Chandran V, Waksman G. 2010. Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. *Biochem J* 425: 475-88
136. Atmakuri K, Cascales E, Christie PJ. 2004. Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Mol Microbiol* 54: 1199-211
137. Baron C, Thorstenson YR, Zambryski PC. 1997. The lipoprotein VirB7 interacts with VirB9 in the membranes of *Agrobacterium tumefaciens*. *J Bacteriol* 179: 1211-8
138. Fernandez D, Dang TA, Spudich GM, Zhou XR, Berger BR, Christie PJ. 1996. The *Agrobacterium tumefaciens* virB7 gene product, a proposed component of the T-

- complex transport apparatus, is a membrane-associated lipoprotein exposed at the periplasmic surface. *J Bacteriol* 178: 3156-67
139. Chandran V, Fronzes R, Duquerroy S, Cronin N, Navaza J, Waksman G. 2009. Structure of the outer membrane complex of a type IV secretion system. *Nature* 462: 1011-5
  140. Lai EM, Kado CI. 1998. Processed VirB2 is the major subunit of the promiscuous pilus of *Agrobacterium tumefaciens*. *J Bacteriol* 180: 2711-7
  141. Schmidt-Eisenlohr H, Domke N, Angerer C, Wanner G, Zambryski PC, Baron C. 1999. Vir proteins stabilize VirB5 and mediate its association with the T pilus of *Agrobacterium tumefaciens*. *J Bacteriol* 181: 7485-92
  142. Fronzes R, Schafer E, Wang L, Saibil HR, Orlova EV, Waksman G. 2009. Structure of a type IV secretion system core complex. *Science* 323: 266-8
  143. Babic A, Lindner AB, Vulic M, Stewart EJ, Radman M. 2008. Direct visualization of horizontal gene transfer. *Science* 319: 1533-6
  144. Backert S, Fronzes R, Waksman G. 2008. VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol* 16: 409-13
  145. Sagulenko E, Sagulenko V, Chen J, Christie PJ. 2001. Role of *Agrobacterium* VirB11 ATPase in T-pilus assembly and substrate selection. *J Bacteriol* 183: 5813-25
  146. Leo JC, Grin I, Linke D. 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos Trans R Soc Lond B Biol Sci* 367: 1088-101
  147. Lyskowski A, Leo JC, Goldman A. 2011. Structure and biology of trimeric autotransporter adhesins. *Adv Exp Med Biol* 715: 143-58
  148. Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VA. 2006. Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol* 14: 264-70
  149. Kirjavainen V, Jarva H, Biedzka-Sarek M, Blom AM, Skurnik M, Meri S. 2008. *Yersinia enterocolitica* serum resistance proteins YadA and ail bind the complement regulator C4b-binding protein. *PLoS Pathog* 4: e1000140
  150. Totsika M, Wells TJ, Beloin C, Valle J, Allsopp LP, King NP, Ghigo JM, Schembri MA. 2012. Molecular characterisation of the EhaG and UpaG trimeric autotransporter proteins from pathogenic *Escherichia coli*. *Appl Environ Microbiol*

151. Cotter SE, Surana NK, Grass S, St Geme JW, 3rd. 2006. Trimeric autotransporters require trimerization of the passenger domain for stability and adhesive activity. *J Bacteriol* 188: 5400-7
152. Meng G, Surana NK, St Geme JW, 3rd, Waksman G. 2006. Structure of the outer membrane translocator domain of the *Haemophilus influenzae* Hia trimeric autotransporter. *Embo J* 25: 2297-304
153. Mikula KM, Leo JC, Lyskowski A, Kedracka-Krok S, Pirog A, Goldman A. 2012. The translocation domain in trimeric autotransporter adhesins is necessary and sufficient for trimerization and autotransportation. *J Bacteriol* 194: 827-38
154. Szczesny P, Lupas A. 2008. Domain annotation of trimeric autotransporter adhesins--daTAA. *Bioinformatics* 24: 1251-6
155. Koretke KK, Szczesny P, Gruber M, Lupas AN. 2006. Model structure of the prototypical non-fimbrial adhesin YadA of *Yersinia enterocolitica*. *J Struct Biol* 155: 154-61
156. Nummelin H, Merckel MC, Leo JC, Lankinen H, Skurnik M, Goldman A. 2004. The *Yersinia* adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *Embo J* 23: 701-11
157. Kajava AV, Steven AC. 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: 306-15
158. Edwards TE, Phan I, Abendroth J, Dieterich SH, Masoudi A, Guo W, Hewitt SN, Kelley A, Leibly D, Brittnacher MJ, Staker BL, Miller SI, Van Voorhis WC, Myler PJ, Stewart LJ. 2010. Structure of a *Burkholderia pseudomallei* trimeric autotransporter adhesin head. *PLoS One* 5
159. Jacob-Dubuisson F, Locht C, Antoine R. 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol Microbiol* 40: 306-13
160. Aoki SK, Pamma R, Hernday AD, Bickham JE, Braaten BA, Low DA. 2005. Contact-dependent inhibition of growth in *Escherichia coli*. *Science* 309: 1245-8
161. Jacob-Dubuisson F, Buisine C, Mielcarek N, Clement E, Menozzi FD, Locht C. 1996. Amino-terminal maturation of the *Bordetella pertussis* filamentous haemagglutinin. *Mol Microbiol* 19: 65-78

162. Grass S, St Geme JW, 3rd. 2000. Maturation and secretion of the non-typable *Haemophilus influenzae* HMW1 adhesin: roles of the N-terminal and C-terminal domains. *Mol Microbiol* 36: 55-67
163. Mazar J, Cotter PA. 2007. New insight into the molecular mechanisms of two-partner secretion. *Trends Microbiol* 15: 508-15
164. Coutte L, Antoine R, Drobecq H, Loch C, Jacob-Dubuisson F. 2001. Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *Embo J* 20: 5040-8
165. Mazar J, Cotter PA. 2006. Topology and maturation of filamentous haemagglutinin suggest a new model for two-partner secretion. *Mol Microbiol* 62: 641-54
166. St Geme JW, 3rd, Yeo HJ. 2009. A prototype two-partner secretion pathway: the *Haemophilus influenzae* HMW1 and HMW2 adhesin systems. *Trends Microbiol* 17: 355-60
167. Buscher AZ, Grass S, Heuser J, Roth R, St Geme JW, 3rd. 2006. Surface anchoring of a bacterial adhesin secreted by the two-partner secretion pathway. *Mol Microbiol* 61: 470-83
168. Jacob-Dubuisson F, Villeret V, Clantin B, Delattre AS, Saint N. 2009. First structural insights into the TpsB/Omp85 superfamily. *Biol Chem* 390: 675-84
169. Clantin B, Delattre AS, Rucktooa P, Saint N, Meli AC, Loch C, Jacob-Dubuisson F, Villeret V. 2007. Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science* 317: 957-61
170. Li H, Grass S, Wang T, Liu T, St Geme JW, 3rd. 2007. Structure of the *Haemophilus influenzae* HMW1B translocator protein: evidence for a twin pore. *J Bacteriol* 189: 7497-502
171. Surana NK, Buscher AZ, Hardy GG, Grass S, Kehl-Fie T, St Geme JW, 3rd. 2006. Translocator proteins in the two-partner secretion family have multiple domains. *J Biol Chem* 281: 18051-8
172. Fan E, Fiedler S, Jacob-Dubuisson F, Muller M. 2012. Two-partner Secretion of Gram-negative Bacteria: A SINGLE beta-BARREL PROTEIN ENABLES TRANSPORT ACROSS THE OUTER MEMBRANE. *J Biol Chem* 287: 2591-9
173. Baud C, Hodak H, Willery E, Drobecq H, Loch C, Jamin M, Jacob-Dubuisson F. 2009. Role of DegP for two-partner secretion in *Bordetella*. *Mol Microbiol* 74: 315-29

174. Salacha R, Kovacic F, Brochier-Armanet C, Wilhelm S, Tommassen J, Filloux A, Voulhoux R, Bleves S. 2010. The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environ Microbiol* 12: 1498-512
175. Tsai JC, Yen MR, Castillo R, Leyton DL, Henderson IR, Saier MH, Jr. 2010. The bacterial intimins and invasins: a large and novel family of secreted proteins. *PLoS One* 5: e14403
176. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ. 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the Dictyostelium host model system. *Proc Natl Acad Sci U S A* 103: 1528-33
177. Boyer F, Fichant G, Berthod J, Vandenbrouck Y, Attree I. 2009. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* 10: 104
178. Hood RD, Singh P, Hsu F, Guvener T, Carl MA, Trinidad RR, Silverman JM, Ohlson BB, Hicks KG, Plemel RL, Li M, Schwarz S, Wang WY, Merz AJ, Goodlett DR, Mougous JD. 2010. A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* 7: 25-37
179. Enos-Berlage JL, Guvener ZT, Keenan CE, McCarter LL. 2005. Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*. *Mol Microbiol* 55: 1160-82
180. Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordonez CL, Lory S, Walz T, Joachimiak A, Mekalanos JJ. 2006. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* 312: 1526-30
181. Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ. 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci U S A* 104: 15508-13
182. Aschtgen MS, Bernard CS, De Bentzmann S, Lloubes R, Cascales E. 2008. SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*. *J Bacteriol* 190: 7523-31
183. Zheng J, Leung KY. 2007. Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol Microbiol* 66: 1192-206

184. Sexton JA, Miller JL, Yoneda A, Kehl-Fie TE, Vogel JP. 2004. *Legionella pneumophila* DotU and IcmF are required for stability of the Dot/Icm complex. *Infect Immun* 72: 5983-92
185. Bonemann G, Pietrosiuk A, Diemand A, Zentgraf H, Mogk A. 2009. Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *Embo J* 28: 315-25
186. Basler M, Pilhofer M, Henderson GP, Jensen GJ, Mekalanos JJ. 2012. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* 483: 182-6
187. Bonemann G, Pietrosiuk A, Mogk A. 2010. Tubules and donuts: a type VI secretion story. *Mol Microbiol* 76: 815-21
188. Whitfield C. 2006. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* 75: 39-68
189. Raetz CR, Whitfield C. 2002. Lipopolysaccharide endotoxins. *Annu Rev Biochem* 71: 635-700
190. Fletcher CM, Coyne MJ, Comstock LE. 2011. Theoretical and experimental characterization of the scope of protein O-glycosylation in *Bacteroides fragilis*. *J Biol Chem* 286: 3219-26
191. Upreti RK, Kumar M, Shankar V. 2003. Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics* 3: 363-79
192. Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* 8: 765-78
193. Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol* 32: 1022-30
194. Scott NE, Parker BL, Connolly AM, Paulech J, Edwards AV, Crossett B, Falconer L, Kolarich D, Djordjevic SP, Hojrup P, Packer NH, Larsen MR, Cordwell SJ. 2011. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol Cell Proteomics* 10: M000031-MCP201

195. Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, Wacker M, Aebi M. 2006. Definition of the bacterial N-glycosylation site consensus sequence. *Embo J* 25: 1957-66
196. Schwarz F, Lizak C, Fan YY, Fleurkens S, Kowarik M, Aebi M. 2011. Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology* 21: 45-54
197. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M. 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298: 1790-3
198. Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, Catrein I, Aebi M. 2006. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* 314: 1148-50
199. Schwarz F, Aebi M. 2011. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* 21: 576-82
200. Nothaft H, Liu X, McNally DJ, Li J, Szymanski CM. 2009. Study of free oligosaccharides derived from the bacterial N-glycosylation pathway. *Proc Natl Acad Sci U S A* 106: 15019-24
201. Nothaft H, Liu X, Li J, Szymanski CM. 2010. *Campylobacter jejuni* free oligosaccharides: function and fate. *Virulence* 1: 546-50
202. Lee S, Cho E, Jung S. 2009. Periplasmic glucans isolated from *Proteobacteria*. *BMB Rep* 42: 769-75
203. Bohin JP. 2000. Osmoregulated periplasmic glucans in *Proteobacteria*. *FEMS Microbiol Lett* 186: 11-9
204. Karlyshev AV, Everest P, Linton D, Cawthraw S, Newell DG, Wren BW. 2004. The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology* 150: 1957-64
205. Mescher MF, Strominger JL. 1976. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *J Biol Chem* 251: 2005-14
206. Calo D, Kaminski L, Eichler J. 2010. Protein glycosylation in Archaea: sweet and extreme. *Glycobiology* 20: 1065-76



207. Grass S, Lichti CF, Townsend RR, Gross J, St Geme JW, 3rd. 2010. The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathog* 6: e1000919
208. Kawai F, Grass S, Kim Y, Choi KJ, St Geme JW, 3rd, Yeo HJ. 2011. Structural insights into the glycosyltransferase activity of the *Actinobacillus pleuropneumoniae* HMW1C-like protein. *J Biol Chem* 286: 38546-57
209. Grass S, Buscher AZ, Swords WE, Apicella MA, Barenkamp SJ, Ozchlewski N, St Geme JW, 3rd. 2003. The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol Microbiol* 48: 737-51
210. Logan SM. 2006. Flagellar glycosylation - a new component of the motility repertoire? *Microbiology* 152: 1249-62
211. Thibault P, Logan SM, Kelly JF, Brisson JR, Ewing CP, Trust TJ, Guerry P. 2001. Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J Biol Chem* 276: 34862-70
212. Goon S, Kelly JF, Logan SM, Ewing CP, Guerry P. 2003. Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the Cj1293 gene. *Mol Microbiol* 50: 659-71
213. Ewing CP, Andreishcheva E, Guerry P. 2009. Functional characterization of flagellin glycosylation in *Campylobacter jejuni* 81-176. *J Bacteriol* 191: 7086-93
214. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AH, Whitehead S, Barrell BG. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403: 665-8
215. Macnab RM. 2003. How bacteria assemble flagella. *Annu Rev Microbiol* 57: 77-100
216. Samatey FA, Imada K, Vonderviszt F, Shirakihara Y, Namba K. 2000. Crystallization of the F41 fragment of flagellin and data collection from extremely thin crystals. *J Struct Biol* 132: 106-11
217. Craig L, Pique ME, Tainer JA. 2004. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* 2: 363-78

218. Aas FE, Vik A, Vedde J, Koomey M, Egge-Jacobsen W. 2007. *Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Mol Microbiol* 65: 607-24
219. Power PM, Roddam LF, Rutter K, Fitzpatrick SZ, Srikhanta YN, Jennings MP. 2003. Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. *Mol Microbiol* 49: 833-47
220. Horzempa J, Comer JE, Davis SA, Castric P. 2006. Glycosylation substrate specificity of *Pseudomonas aeruginosa* 1244 pilin. *J Biol Chem* 281: 1128-36
221. Kus JV, Kelly J, Tessier L, Harvey H, Cvitkovitch DG, Burrows LL. 2008. Modification of *Pseudomonas aeruginosa* Pa5196 type IV Pilins at multiple sites with D-Araf by a novel GT-C family Arabinosyltransferase, TfpW. *J Bacteriol* 190: 7464-78
222. Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, Tainer JA. 1995. Structure of the fibre-forming protein pilin at 2.6 Å resolution. *Nature* 378: 32-8
223. Hegge FT, Hitchen PG, Aas FE, Kristiansen H, Lovold C, Egge-Jacobsen W, Panico M, Leong WY, Bull V, Virji M, Morris HR, Dell A, Koomey M. 2004. Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of *Neisseria gonorrhoeae* type IV pili. *Proc Natl Acad Sci USA* 101: 10798-803
224. Anonsen JH, Egge-Jacobsen W, Aas FE, Borud B, Koomey M, Vik A. 2012. Novel protein substrates of the phospho-form modification system in *Neisseria gonorrhoeae* and their connection to O-linked protein glycosylation. *Infect Immun* 80: 22-30
225. Smedley JG, 3rd, Jewell E, Roguskie J, Horzempa J, Syboldt A, Stolz DB, Castric P. 2005. Influence of pilin glycosylation on *Pseudomonas aeruginosa* 1244 pilus function. *Infect Immun* 73: 7922-31
226. Marceau M, Forest K, Beretti JL, Tainer J, Nassif X. 1998. Consequences of the loss of O-linked glycosylation of meningococcal type IV pilin on piliation and pilus-mediated adhesion. *Mol Microbiol* 27: 705-15
227. Jennings MP, Jen FE, Roddam LF, Apicella MA, Edwards JL. 2011. *Neisseria gonorrhoeae* pilin glycan contributes to CR3 activation during challenge of primary cervical epithelial cells. *Cell Microbiol* 13: 885-96

228. Easton DM, Totsika M, Allsopp LP, Phan MD, Idris A, Wurpel DJ, Sherlock O, Zhang B, Venturini C, Beatson SA, Mahony TJ, Cobbold RN, Schembri MA. 2011. Characterization of EhaJ, a New Autotransporter Protein from Enterohemorrhagic and Enteropathogenic *Escherichia coli*. *Front Microbiol* 2: 120
229. Reidl S, Lehmann A, Schiller R, Salam Khan A, Dobrindt U. 2009. Impact of O-glycosylation on the molecular and cellular adhesion properties of the *Escherichia coli* autotransporter protein Ag43. *Int J Med Microbiol* 299: 389-401
230. Moormann C, Benz I, Schmidt MA. 2002. Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infect Immun* 70: 2264-70
231. Sherlock O, Schembri MA, Reisner A, Klemm P. 2004. Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm formation. *J Bacteriol* 186: 8058-65
232. Sherlock O, Vejborg RM, Klemm P. 2005. The TibA adhesin/invasin from enterotoxigenic *Escherichia coli* is self recognizing and induces bacterial aggregation and biofilm formation. *Infect Immun* 73: 1954-63
233. Knudsen SK, Stensballe A, Franzmann M, Westergaard UB, Otzen DE. 2008. Effect of glycosylation on the extracellular domain of the Ag43 bacterial autotransporter: enhanced stability and reduced cellular aggregation. *Biochem J* 412: 563-77
234. Ku SC, Schulz BL, Power PM, Jennings MP. 2009. The pilin O-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. *Biochem Biophys Res Commun* 378: 84-9
235. Vik A, Aas FE, Anonsen JH, Bilsborough S, Schneider A, Egge-Jacobsen W, Koomey M. 2009. Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc Natl Acad Sci U S A* 106: 4447-52
236. Comstock LE. 2009. Importance of glycans to the host-bacteroides mutualism in the mammalian intestine. *Cell Host Microbe* 5: 522-6
237. Coyne MJ, Reinap B, Lee MM, Comstock LE. 2005. Human symbionts use a host-like pathway for surface fucosylation. *Science* 307: 1778-81
238. Fletcher CM, Coyne MJ, Villa OF, Chatzidaki-Livanis M, Comstock LE. 2009. A general O-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell* 137: 321-31

239. Plummer TH, Jr., Tarentino AL, Hauer CR. 1995. Novel, specific O-glycosylation of secreted *Flavobacterium meningosepticum* proteins. Asp-Ser and Asp-Thr-Thr consensus sites. *J Biol Chem* 270: 13192-6
240. Schmidt MA, Riley LW, Benz I. 2003. Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol* 11: 554-61
241. Wehmeier S, Varghese AS, Gurcha SS, Tissot B, Panico M, Hitchen P, Morris HR, Besra GS, Dell A, Smith MC. 2009. Glycosylation of the phosphate binding protein, PstS, in *Streptomyces coelicolor* by a pathway that resembles protein O-mannosylation in eukaryotes. *Mol Microbiol* 71: 421-33
242. Lommel M, Strahl S. 2009. Protein O-mannosylation: conserved from bacteria to humans. *Glycobiology* 19: 816-28
243. VanderVen BC, Harder JD, Crick DC, Belisle JT. 2005. Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science* 309: 941-3
244. Espitia C, Servin-Gonzalez L, Mancilla R. 2010. New insights into protein O-mannosylation in *actinomycetes*. *Mol Biosyst* 6: 775-81
245. Strahl-Bolsinger S, Gentzsch M, Tanner W. 1999. Protein O-mannosylation. *Biochim Biophys Acta* 1426: 297-307
246. Zhou M, Wu H. 2009. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology* 155: 317-27
247. Wu H, Mintz KP, Ladha M, Fives-Taylor PM. 1998. Isolation and characterization of Fap1, a fimbriae-associated adhesin of *Streptococcus parasanguis* FW213. *Mol Microbiol* 28: 487-500
248. Bensing BA, Gibson BW, Sullam PM. 2004. The *Streptococcus gordonii* platelet binding protein GspB undergoes glycosylation independently of export. *J Bacteriol* 186: 638-45
249. Plummer C, Wu H, Kerrigan SW, Meade G, Cox D, Ian Douglas CW. 2005. A serine-rich glycoprotein of *Streptococcus sanguis* mediates adhesion to platelets via GPIb. *Br J Haematol* 129: 101-9
250. Siboo IR, Chaffin DO, Rubens CE, Sullam PM. 2008. Characterization of the accessory Sec system of *Staphylococcus aureus*. *J Bacteriol* 190: 6188-96
251. Chen Q, Wu H, Fives-Taylor PM. 2004. Investigating the role of secA2 in secretion and glycosylation of a fimbrial adhesin in *Streptococcus parasanguis* FW213. *Mol Microbiol* 53: 843-56

252. Bu S, Li Y, Zhou M, Azadin P, Zeng M, Fives-Taylor P, Wu H. 2008. Interaction between two putative glycosyltransferases is required for glycosylation of a serine-rich streptococcal adhesin. *J Bacteriol* 190: 1256-66
253. Stephenson AE, Wu H, Novak J, Tomana M, Mintz K, Fives-Taylor P. 2002. The Fap1 fimbrial adhesin is a glycoprotein: antibodies specific for the glycan moiety block the adhesion of *Streptococcus parasanguis* in an in vitro tooth model. *Mol Microbiol* 43: 147-57
254. Rigel NW, Braunstein M. 2008. A new twist on an old pathway--accessory Sec systems. *Mol Microbiol* 69: 291-302
255. Bensing BA, Takamatsu D, Sullam PM. 2005. Determinants of the streptococcal surface glycoprotein GspB that facilitate export by the accessory Sec system. *Mol Microbiol* 58: 1468-81
256. Chen Q, Sun B, Wu H, Peng Z, Fives-Taylor PM. 2007. Differential roles of individual domains in selection of secretion route of a *Streptococcus parasanguinis* serine-rich adhesin, Fap1. *J Bacteriol* 189: 7610-7
257. Jones C. 2005. Vaccines based on the cell surface carbohydrates of pathogenic bacteria. *An Acad Bras Cienc* 77: 293-324
258. Pandhal J, Wright PC. 2010. N-Linked glycoengineering for human therapeutic proteins in bacteria. *Biotechnol Lett* 32: 1189-98
259. Schwarz F, Huang W, Li C, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang LX. 2010. A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat Chem Biol* 6: 264-6
260. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, Heiss C, Azadi P, Aebi M, Delisa MP. 2012. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat Chem Biol*
261. Khan S, Mian HS, Sandercock LE, Chirgadze NY, Pai EF. 2011. Crystal structure of the passenger domain of the *Escherichia coli* autotransporter EspP. *J Mol Biol* 413: 985-1000
262. Meng G, Spahich N, Kenjale R, Waksman G, St Geme JW, 3rd. 2011. Crystal structure of the *Haemophilus influenzae* Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. *Embo J* 30: 3864-74
263. van den Berg B. 2010. Crystal structure of a full-length autotransporter. *J Mol Biol* 396: 627-33

264. Junker M, Schuster CC, McDonnell AV, Sorg KA, Finn MC, Berger B, Clark PL. 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: 4918-23
265. Emsley P, Charles IG, Fairweather NF, Isaacs NW. 1996. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature* 381: 90-2
266. Otto BR, Sijbrandi R, Luirink J, Oudega B, Heddle JG, Mizutani K, Park SY, Tame JR. 2005. Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic *Escherichia coli*. *J Biol Chem* 280: 17339-45
267. Johnson TA, Qiu J, Plaut AG, Holyoak T. 2009. Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of *haemophilus influenzae* immunoglobulin A1 protease. *J Mol Biol* 389: 559-74
268. Benz I, Schmidt MA. 2011. Structures and functions of autotransporter proteins in microbial pathogens. *Int J Med Microbiol* 301: 461-8
269. Girard V, Mourez M. 2006. Adhesion mediated by autotransporters of Gram-negative bacteria: structural and functional features. *Res Microbiol* 157: 407-16
270. Yen YT, Kostakioti M, Henderson IR, Stathopoulos C. 2008. Common themes and variations in serine protease autotransporters. *Trends Microbiol* 16: 370-9
271. Barnes MG, Weiss AA. 2001. BrkA protein of *Bordetella pertussis* inhibits the classical pathway of complement after C1 deposition. *Infect Immun* 69: 3067-72
272. Goldberg MB, Theriot JA. 1995. *Shigella flexneri* surface protein IcsA is sufficient to direct actin-based motility. *Proc Natl Acad Sci U S A* 92: 6572-6
273. Timpe JM, Holm MM, Vanlerberg SL, Basrur V, Lafontaine ER. 2003. Identification of a *Moraxella catarrhalis* outer membrane protein exhibiting both adhesin and lipolytic activities. *Infect Immun* 71: 4341-50
274. Cover TL, Blanke SR. 2005. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol* 3: 320-32
275. Oomen CJ, van Ulsen P, van Gelder P, Feijen M, Tommassen J, Gros P. 2004. Structure of the translocator domain of a bacterial autotransporter. *Embo J* 23: 1257-66
276. Zhai Y, Zhang K, Huo Y, Zhu Y, Zhou Q, Lu J, Black I, Pang X, Roszak AW, Zhang X, Isaacs NW, Sun F. 2011. Autotransporter passenger domain secretion

- requires a hydrophobic cavity at the extracellular entrance of the beta-domain pore. *Biochem J* 435: 577-87
277. Barnard TJ, Dautin N, Lukacik P, Bernstein HD, Buchanan SK. 2007. Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. *Nat Struct Mol Biol* 14: 1214-20
278. Tajima N, Kawai F, Park SY, Tame JR. 2010. A novel intein-like autoproteolytic mechanism in autotransporter proteins. *J Mol Biol* 402: 645-56
279. Khalid S, Sansom MS. 2006. Molecular dynamics simulations of a bacterial autotransporter: NalP from *Neisseria meningitidis*. *Mol Membr Biol* 23: 499-508
280. Nishimura K, Tajima N, Yoon YH, Park SY, Tame JR. 2010. Autotransporter passenger proteins: virulence factors with common structural themes. *J Mol Med (Berl)* 88: 451-8
281. Kuhnel K, Diezmann D. 2011. Crystal structure of the autochaperone region from the *Shigella flexneri* autotransporter IcsA. *J Bacteriol* 193: 2042-5
282. Oliver DC, Huang G, Nodel E, Pleasance S, Fernandez RC. 2003. A conserved region within the *Bordetella pertussis* autotransporter BrkA is necessary for folding of its passenger domain. *Mol Microbiol* 47: 1367-83
283. Dutta PR, Sui BQ, Nataro JP. 2003. Structure-function analysis of the enteroaggregative *Escherichia coli* plasmid-encoded toxin autotransporter using scanning linker mutagenesis. *J Biol Chem* 278: 39912-20
284. Konieczny MPJ, Benz I, Hollinderbaumer B, Beinke C, Niederweis M, Schmidt MA. 2001. Modular organization of the AIDA autotransporter translocator: the N-terminal beta1-domain is surface-exposed and stabilizes the transmembrane beta2-domain. *Antonie Van Leeuwenhoek* 80: 19-34
285. May KL, 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: 4666-76
286. Renn JP, Clark PL. 2008. A conserved stable core structure in the passenger domain beta-helix of autotransporter virulence proteins. *Biopolymers* 89: 420-7
287. Sijbrandi R, Urbanus ML, ten Hagen-Jongman CM, Bernstein HD, Oudega B, Otto BR, Luirink J. 2003. Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular *Escherichia coli* protein. *J Biol Chem* 278: 4654-9

288. Desvaux M, Scott-Tucker A, Turner SM, Cooper LM, Huber D, Nataro JP, Henderson IR. 2007. A conserved extended signal peptide region directs posttranslational protein translocation via a novel mechanism. *Microbiology* 153: 59-70
289. Peterson JH, Szabady RL, Bernstein HD. 2006. An unusual signal peptide extension inhibits the binding of bacterial presecretory proteins to the signal recognition particle, trigger factor, and the SecYEG complex. *J Biol Chem* 281: 9038-48
290. Jong WS, Luirink J. 2008. The conserved extension of the Hbp autotransporter signal peptide does not determine targeting pathway specificity. *Biochem Biophys Res Commun* 368: 522-7
291. Szabady RL, Peterson JH, Skillman KM, Bernstein HD. 2005. An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proc Natl Acad Sci U S A* 102: 221-6
292. Leyton DL, de Luna MG, Sevastyanovich YR, Tveen Jensen K, Browning DF, Scott-Tucker A, Henderson IR. 2010. The unusual extended signal peptide region is not required for secretion and function of an *Escherichia coli* autotransporter. *FEMS Microbiol Lett* 311: 133-9
293. Jong WS, ten Hagen-Jongman CM, Ruijter E, Orru RV, Genevaux P, Luirink J. 2010. YidC is involved in the biogenesis of the secreted autotransporter hemoglobin protease. *J Biol Chem* 285: 39682-90
294. Brandon LD, Goehring N, Janakiraman A, Yan AW, Wu T, Beckwith J, Goldberg MB. 2003. IcsA, a polarly localized autotransporter with an atypical signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed. *Mol Microbiol* 50: 45-60
295. Ieva R, Tian P, Peterson JH, Bernstein HD. 2011. Sequential and spatially restricted interactions of assembly factors with an autotransporter beta domain. *Proc Natl Acad Sci U S A* 108: E383-91
296. Ruiz-Perez F, Henderson IR, Leyton DL, Rossiter AE, Zhang Y, Nataro JP. 2009. Roles of periplasmic chaperone proteins in the biogenesis of serine protease autotransporters of *Enterobacteriaceae*. *J Bacteriol* 191: 6571-83
297. Ruiz-Perez F, Henderson IR, Nataro JP. 2010. Interaction of FkpA, a peptidyl-prolyl cis/trans isomerase with EspP autotransporter protein. *Gut Microbes* 1: 339-44



298. Sauri A, Soprova Z, Wickstrom D, de Gier JW, Van der Schors RC, Smit AB, Jong WS, Luirink J. 2009. The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. *Microbiology* 155: 3982-91
299. Wagner JK, Heindl JE, Gray AN, Jain S, Goldberg MB. 2009. Contribution of the periplasmic chaperone Skp to efficient presentation of the autotransporter IcsA on the surface of *Shigella flexneri*. *J Bacteriol* 191: 815-21
300. Purdy GE, Fisher CR, Payne SM. 2007. IcsA surface presentation in *Shigella flexneri* requires the periplasmic chaperones DegP, Skp, and SurA. *J Bacteriol* 189: 5566-73
301. Mogensen JE, Kleinschmidt JH, Schmidt MA, Otzen DE. 2005. Misfolding of a bacterial autotransporter. *Protein Sci* 14: 2814-27
302. Hodak H, Wohlkonig A, Smet-Nocca C, Drobecq H, Wieruszkeski JM, Senechal M, Landrieu I, Locht C, Jamin M, Jacob-Dubuisson F. 2008. The peptidyl-prolyl isomerase and chaperone Par27 of *Bordetella pertussis* as the prototype for a new group of parvulins. *J Mol Biol* 376: 414-26
303. Volokhina EB, Grijpstra J, Stork M, Schilders I, Tommassen J, Bos MP. 2011. Role of the periplasmic chaperones Skp, SurA, and DegQ in outer membrane protein biogenesis in *Neisseria meningitidis*. *J Bacteriol* 193: 1612-21
304. Ieva R, Bernstein HD. 2009. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc Natl Acad Sci U S A* 106: 19120-5
305. Jain S, Goldberg MB. 2007. Requirement for YaeT in the outer membrane assembly of autotransporter proteins. *J Bacteriol* 189: 5393-8
306. Rossiter AE, Leyton DL, Tveen-Jensen K, Browning DF, Sevastyanovich Y, Knowles TJ, Nichols KB, Cunningham AF, Overduin M, Schembri MA, Henderson IR. 2011. The essential beta-barrel assembly machinery complex components BamD and BamA are required for autotransporter biogenesis. *J Bacteriol* 193: 4250-3
307. Sauri A, Ten Hagen-Jongman CM, van Ulsen P, Luirink J. 2012. Estimating the size of the active translocation pore of an autotransporter. *J Mol Biol* 416: 335-45
308. Ieva R, Skillman KM, Bernstein HD. 2008. Incorporation of a polypeptide segment into the beta-domain pore during the assembly of a bacterial autotransporter. *Mol Microbiol* 67: 188-201

309. Junker M, Besingi RN, Clark PL. 2009. Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Mol Microbiol* 71: 1323-32
310. Jong WS, ten Hagen-Jongman CM, den Blaauwen T, Slotboom DJ, Tame JR, Wickstrom D, de Gier JW, Otto BR, Luirink J. 2007. Limited tolerance towards folded elements during secretion of the autotransporter Hbp. *Mol Microbiol* 63: 1524-36
311. Leyton DL, Sevastyanovich YR, Browning DF, Rossiter AE, Wells TJ, Fitzpatrick RE, Overduin M, Cunningham AF, Henderson IR. 2011. Size and conformation limits to secretion of disulfide-bonded loops in autotransporter proteins. *J Biol Chem* 286: 42283-91
312. Veiga E, de Lorenzo V, Fernandez LA. 2004. Structural tolerance of bacterial autotransporters for folded passenger protein domains. *Mol Microbiol* 52: 1069-80
313. Skillman KM, Barnard TJ, Peterson JH, Ghirlando R, Bernstein HD. 2005. Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter. *Mol Microbiol* 58: 945-58
314. Phan G, Remaut H, Wang T, Allen WJ, Pirker KF, Lebedev A, Henderson NS, Geibel S, Volkan E, Yan J, Kunze MB, Pinkner JS, Ford B, Kay CW, Li H, Hultgren SJ, Thanassi DG, Waksman G. 2011. Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature* 474: 49-53
315. Renn JP, Junker M, Besingi RN, Braselmann E, Clark PL. 2012. ATP-Independent Control of Autotransporter Virulence Protein Transport via the Folding Properties of the Secreted Protein. *Chem Biol* 19: 287-96
316. Soprova Z, Sauri A, van Ulsen P, Tame JR, den Blaauwen T, Jong WS, Luirink J. 2010. A conserved aromatic residue in the autochaperone domain of the autotransporter Hbp is critical for initiation of outer membrane translocation. *J Biol Chem* 285: 38224-33
317. Selkrig J, Mosbahi K, Webb CT, Belousoff MJ, Perry AJ, Wells TJ, Morris F, Leyton DL, Totsika M, Phan MD, Celik N, Kelly M, Oates C, Hartland EL, Robins-Browne RM, Ramarathinam SH, Purcell AW, Schembri MA, Strugnell RA, Henderson IR, Walker D, Lithgow T. 2012. Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat Struct Mol Biol*

318. Shere KD, Sallustio S, Manassis A, D'Aversa TG, Goldberg MB. 1997. Disruption of IcsP, the major *Shigella* protease that cleaves IcsA, accelerates actin-based motility. *Mol Microbiol* 25: 451-62
319. Turner DP, Marietou AG, Johnston L, Ho KK, Rogers AJ, Wooldridge KG, Ala'Aldeen DA. 2006. Characterization of MspA, an immunogenic autotransporter protein that mediates adhesion to epithelial and endothelial cells in *Neisseria meningitidis*. *Infect Immun* 74: 2957-64
320. van Ulsen P, Adler B, Fassler P, Gilbert M, van Schilfgaarde M, van der Ley P, van Alphen L, Tommassen J. 2006. A novel phase-variable autotransporter serine protease, AusI, of *Neisseria meningitidis*. *Microbes Infect* 8: 2088-97
321. van Ulsen P, van Alphen L, ten Hove J, Fransen F, van der Ley P, Tommassen J. 2003. A Neisserial autotransporter NalP modulating the processing of other autotransporters. *Mol Microbiol* 50: 1017-30
322. Fink DL, Cope LD, Hansen EJ, Geme JW, 3rd. 2001. The *Hemophilus influenzae* Hap autotransporter is a chymotrypsin clan serine protease and undergoes autoproteolysis via an intermolecular mechanism. *J Biol Chem* 276: 39492-500
323. Dautin N, Barnard TJ, Anderson DE, Bernstein HD. 2007. Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. *Embo J* 26: 1942-52
324. Nguyen VQ, Caprioli RM, Cover TL. 2001. Carboxy-terminal proteolytic processing of *Helicobacter pylori* vacuolating toxin. *Infect Immun* 69: 543-6
325. Henderson IR, Owen P. 1999. The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and oxyR. *J Bacteriol* 181: 2132-41
326. Suhr M, Benz I, Schmidt MA. 1996. Processing of the AIDA-I precursor: removal of AIDAc and evidence for the outer membrane anchoring as a beta-barrel structure. *Mol Microbiol* 22: 31-42
327. Vandahl BB, Pedersen AS, Gevaert K, Holm A, Vandekerckhove J, Christiansen G, Birkelund S. 2002. The expression, processing and localization of polymorphic membrane proteins in *Chlamydia pneumoniae* strain CWL029. *BMC Microbiol* 2: 36

328. Navarro-Garcia F. 2010. Enteroaggregative *Escherichia coli* plasmid-encoded toxin. *Future Microbiol* 5: 1005-13
329. Goldberg MB, Barzu O, Parsot C, Sansonetti PJ. 1993. Unipolar localization and ATPase activity of IcsA, a *Shigella flexneri* protein involved in intracellular movement. *J Bacteriol* 175: 2189-96
330. d'Hauteville H, Dufourcq Lagelouse R, Nato F, Sansonetti PJ. 1996. Lack of cleavage of IcsA in *Shigella flexneri* causes aberrant movement and allows demonstration of a cross-reactive eukaryotic protein. *Infect Immun* 64: 511-7
331. Fink DL, St Geme JW, 3rd. 2003. Chromosomal expression of the *Haemophilus influenzae* Hap autotransporter allows fine-tuned regulation of adhesive potential via inhibition of intermolecular autoproteolysis. *J Bacteriol* 185: 1608-15
332. Charles M, Perez M, Kobil JH, Goldberg MB. 2001. Polar targeting of *Shigella* virulence factor IcsA in *Enterobacteriaceae* and *Vibrio*. *Proc Natl Acad Sci U S A* 98: 9871-6
333. Fixen KR, Janakiraman A, Garrity S, Slade DJ, Gray AN, Karahan N, Hochschild A, Goldberg MB. 2012. Genetic reporter system for positioning of proteins at the bacterial pole. *MBio* 3
334. Janakiraman A, Fixen KR, Gray AN, Niki H, Goldberg MB. 2009. A genome-scale proteomic screen identifies a role for DnaK in chaperoning of polar autotransporters in *Shigella*. *J Bacteriol* 191: 6300-11
335. Jain S, van Ulsen P, Benz I, Schmidt MA, Fernandez R, Tommassen J, Goldberg MB. 2006. Polar localization of the autotransporter family of large bacterial virulence proteins. *J Bacteriol* 188: 4841-50
336. Benjelloun-Touimi Z, Si Tahar M, Montecucco C, Sansonetti PJ, Parsot C. 1998. SepA, the 110 kDa protein secreted by *Shigella flexneri*: two-domain structure and proteolytic activity. *Microbiology* 144 ( Pt 7): 1815-22
337. Fernandez RC, Weiss AA. 1994. Cloning and sequencing of a *Bordetella pertussis* serum resistance locus. *Infect Immun* 62: 4727-38
338. Benz I, Schmidt MA. 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infect Immun* 57: 1506-11

339. Odenbreit S, Till M, Hofreuter D, Faller G, Haas R. 1999. Genetic and functional characterization of the alpAB gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Mol Microbiol* 31: 1537-48
340. Coutte L, Willery E, Antoine R, Drobecq H, Loch C, Jacob-Dubuisson F. 2003. Surface anchoring of bacterial subtilisin important for maturation function. *Mol Microbiol* 49: 529-39
341. Jong WS, Sauri A, Luirink J. 2010. Extracellular production of recombinant proteins using bacterial autotransporters. *Curr Opin Biotechnol* 21: 646-52
342. Jose J, Meyer TF. 2007. The autodisplay story, from discovery to biotechnical and biomedical applications. *Microbiol Mol Biol Rev* 71: 600-19
343. Cutter D, Mason KW, Howell AP, Fink DL, Green BA, St Geme JW, 3rd. 2002. Immunization with *Haemophilus influenzae* Hap adhesin protects against nasopharyngeal colonization in experimental mice. *J Infect Dis* 186: 1115-21
344. Wells TJ, Totsika M, Schembri MA. 2010. Autotransporters of *Escherichia coli*: a sequence-based characterization. *Microbiology* 156: 2459-69
345. Benz I, Schmidt MA. 1992. Isolation and serologic characterization of AIDA-I, the adhesin mediating the diffuse adherence phenotype of the diarrhea-associated *Escherichia coli* strain 2787 (O126:H27). *Infect Immun* 60: 13-8
346. Jallat C, Livrelli V, Darfeuille-Michaud A, Rich C, Joly B. 1993. *Escherichia coli* strains involved in diarrhea in France: high prevalence and heterogeneity of diffusely adhering strains. *J Clin Microbiol* 31: 2031-7
347. Jallat C, Darfeuille-Michaud A, Rich C, Joly B. 1994. Survey of clinical isolates of diarrhoeogenic *Escherichia coli*: diffusely adhering *E. coli* strains with multiple adhesive factors. *Res Microbiol* 145: 621-32
348. Niewerth U, Frey A, Voss T, Le Bouguenec C, Baljer G, Franke S, Schmidt MA. 2001. The AIDA autotransporter system is associated with F18 and stx2e in *Escherichia coli* isolates from pigs diagnosed with edema disease and postweaning diarrhea. *Clin Diagn Lab Immunol* 8: 143-9
349. Mainil JG, Jacquemin E, Pohl P, Kaeckenbeeck A, Benz I. 2002. DNA sequences coding for the F18 fimbriae and AIDA adhesin are localised on the same plasmid in *Escherichia coli* isolates from piglets. *Vet Microbiol* 86: 303-11

350. Ha SK, Choi C, Chae C. 2003. Prevalence of a gene encoding adhesin involved in diffuse adherence among *Escherichia coli* isolates in pigs with postweaning diarrhea or edema disease. *J Vet Diagn Invest* 15: 378-81
351. Ngeleka M, Pritchard J, Appleyard G, Middleton DM, Fairbrother JM. 2003. Isolation and association of *Escherichia coli* AIDA-I/STb, rather than EAST1 pathotype, with diarrhea in piglets and antibiotic sensitivity of isolates. *J Vet Diagn Invest* 15: 242-52
352. Zhang W, Zhao M, Ruesch L, Omot A, Francis D. 2007. Prevalence of virulence genes in *Escherichia coli* strains recently isolated from young pigs with diarrhea in the US. *Vet Microbiol* 123: 145-52
353. Zhao L, Chen X, Xu X, Song G, Liu X. 2009. Analysis of the AIDA-I gene sequence and prevalence in *Escherichia coli* isolates from pigs with post-weaning diarrhoea and oedema disease. *Vet J* 180: 124-9
354. Ravi M, Ngeleka M, Kim SH, Gyles C, Berthiaume F, Mourez M, Middleton D, Simko E. 2007. Contribution of AIDA-I to the pathogenicity of a porcine diarrheagenic *Escherichia coli* and to intestinal colonization through biofilm formation in pigs. *Vet Microbiol* 120: 308-19
355. Fang Y, Ngeleka M, Middleton DM, Simko E. 2005. Characterization and immunodetection of AIDA-I adhesin isolated from porcine *Escherichia coli*. *Vet Microbiol* 109: 65-73
356. Dubreuil JD. 2010. STb and AIDA-I: the missing link? *Crit Rev Microbiol* 36: 212-20
357. Benz I, Schmidt MA. 1992. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. *Mol Microbiol* 6: 1539-46
358. Berthiaume F, Rutherford N, Mourez M. 2007. Mutations affecting the biogenesis of the AIDA-I autotransporter. *Res Microbiol* 158: 348-54
359. Laarmann S, Schmidt MA. 2003. The *Escherichia coli* AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor. *Microbiology* 149: 1871-82
360. Fang Y, Ngeleka M, Middleton DM, Simko E. 2008. Isolation and identification of AIDA-I receptors in porcine intestinal mucus. *Vet Microbiol* 126: 345-55

361. Girard V, Cote JP, Charbonneau ME, Campos M, Berthiaume F, Hancock MA, Siddiqui N, Mourez M. 2010. Conformation change in a self-recognizing autotransporter modulates bacterial cell-cell interaction. *J Biol Chem* 285: 10616-26
362. Grizot S, Salem M, Vongsouthi V, Durand L, Moreau F, Dohi H, Vincent S, Escaich S, Ducruix A. 2006. Structure of the *Escherichia coli* heptosyltransferase WaaC: binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose. *J Mol Biol* 363: 383-94
363. Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* 328: 307-17
364. Kadrmas JL, Raetz CR. 1998. Enzymatic synthesis of lipopolysaccharide in *Escherichia coli*. Purification and properties of heptosyltransferase i. *J Biol Chem* 273: 2799-807
365. Gronow S, Brabetz W, Lindner B, Brade H. 2005. OpsX from *Haemophilus influenzae* represents a novel type of heptosyltransferase I in lipopolysaccharide biosynthesis. *J Bacteriol* 187: 6242-7
366. Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S. 2011. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* 469: 564-7
367. Bhushan S, Gartmann M, Halic M, Armache JP, Jarasch A, Mielke T, Berninghausen O, Wilson DN, Beckmann R. 2010. alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* 17: 313-7
368. Kosolapov A, Deutsch C. 2009. Tertiary interactions within the ribosomal exit tunnel. *Nat Struct Mol Biol* 16: 405-11
369. Tu LW, Deutsch C. 2010. A folding zone in the ribosomal exit tunnel for Kv1.3 helix formation. *J Mol Biol* 396: 1346-60
370. Woolhead CA, McCormick PJ, Johnson AE. 2004. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* 116: 725-36
371. Sabate R, de Groot NS, Ventura S. 2010. Protein folding and aggregation in bacteria. *Cell Mol Life Sci* 67: 2695-715
372. Otzen D, Nielsen PH. 2008. We find them here, we find them there: functional bacterial amyloid. *Cell Mol Life Sci* 65: 910-27

373. Shental-Bechor D, Levy Y. 2008. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* 105: 8256-61
374. Evans ML, Schmidt JC, Ilbert M, Doyle SM, Quan S, Bardwell JC, Jakob U, Wickner S, Chapman MR. 2011. E. coli chaperones DnaK, Hsp33 and Spy inhibit bacterial functional amyloid assembly. *Prion* 5
375. Dautin N, Bernstein HD. 2007. Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu Rev Microbiol* 61: 89-112
376. Laarmann S, Cutter D, Juehne T, Barenkamp SJ, St Geme JW. 2002. The *Haemophilus influenzae* Hia autotransporter harbours two adhesive pockets that reside in the passenger domain and recognize the same host cell receptor. *Mol Microbiol* 46: 731-43
377. Klemm P, Hjerrild L, Gjermansen M, Schembri MA. 2004. Structure-function analysis of the self-recognizing Antigen 43 autotransporter protein from *Escherichia coli*. *Mol Microbiol* 51: 283-96
378. Cote JP, Mourez M. 2011. Structure-function analysis of the TibA self-associating autotransporter reveals a modular organization. *Infect Immun* 79: 1826-32
379. Bingel-Erlenmeyer R, Kohler R, Kramer G, Sandikci A, Antolic S, Maier T, Schaffitzel C, Wiedmann B, Bukau B, Ban N. 2008. A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. *Nature* 452: 108-11
380. Giglione C, Fieulaine S, Meinnel T. 2009. Cotranslational processing mechanisms: towards a dynamic 3D model. *Trends Biochem Sci* 34: 417-26
381. Tang G, Mintz KP. 2010. Glycosylation of the collagen adhesin EmaA of *Aggregatibacter actinomycetemcomitans* is dependent upon the lipopolysaccharide biosynthetic pathway. *J Bacteriol* 192: 1395-404



