ROMA TRE
UNIVERSITÀ DEGLI STUDI

DOCTORAL SCHOOL IN BIOMEDICAL SCIENCE AND TECHNOLOGY

XXX DOCTORAL PROGRAM

Proteomic identification of novel adhesins in enterohaemorrhagic *E. coli* O157:H7 in a strategy for preventing intestinal colonisation

Candidate

Ricardo Monteiro

Coordination

Dr. Mickaël Desvaux

Dr. Alfredo Pezzicoli

Prof. Paolo Visca

ROMA TRE
UNIVERSITÀ DEGLI STUDI

DOCTORAL SCHOOL IN BIOMEDICAL SCIENCE AND TECHNOLOGY

XXX DOCTORAL PROGRAM

Proteomic identification of novel adhesins in enterohaemorrhagic *E. coli* O157:H7 in a strategy for preventing intestinal colonisation

Ricardo Monteiro
Candidate

Mickaël Desvaux
Supervisor/INRA

Alfredo Pezzicoli
Supervisor/GSK

Paolo Visca
Supervisor/UNIROMA3

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I

Chapter III

Chapter IV

Chapter V

LIST OF TABLES

Chapter I

Chapter III

Chapter IV

ABBREVIATIONS

AAF: aggregative adherence fimbriae
Aap: anti-aggregation protein
CAN: acetonitrile
AEEC: A/E encoding *E. coli*
AIDA: adhesin diffuse adherence
BFP: bundle-forming pili
BHI: brain hearth infusion
BSA: bovine serum albumin
Cah: calcium-binding antigen 43 homologue
CF: colonisation factor
CFU: colony forming unit
CNF-1: cytotoxic necrotizing factor 1
CU: chaperone-usher
CY: cytoplasm
DAEC: diffusely adherent *Escherichia coli*
DMEM: Dulbecco Modified Eagle Medium
DNA: deoxyribonucleic acid
DTT: Dithiothreitol
Eae: enterocyte *attaching* and *effacing*
EAEC: enteroaggregative *Escherichia coli*
EAF: EPEC adherence factor
EC: extracellular
ECM: extracellular matrix
EDTA: ethylenediaminetetraacetic acid
Efb: extracellular fibrinogen-binding protein
Eha: *E. coli* autotransporters
EHEC: enterohemorrhagic *Escherichia coli*
EIEC: enteroinvasive *Escherichia coli*
ELISA: enzyme-linked immunosorbent assay
ENP: extracellular nucleation-precipitation
EPEC: enteropathogenic *Escherichia coli*
Esp: extracellular serine protease
ETEC: enterotoxigenic *Escherichia coli*
ExPEC: extraintestinal pathogenic *Escherichia coli*
FdeC: Factor adherence *E. coli*
FDR: false discovery rate
FH: factor H
GAPDH: glyceraldehyde 3-phosphate dehydrogenase
GO: gene ontology
HC: hemorrhagic colitis
HCP: haemorrhagic coli pili
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HUS: hemolytic uremic syndrome
Ig: immunoglobulin
IgA-BP: immunoglobulin A-binding protein
IL: interleukin
IM: inner membrane
IML: inner membrane lipoprotein.
IMP: integral membrane protein
InPEC: intestinal pathogenic *Escherichia coli*
IVIS: In Vivo Imaging System
LB: Lysogeny-broth
LEE: locus of enterocyte effacement
Lpf: long polar fimbriae
LPS: lipopolysaccharide
LT: heat-labile enterotoxin
MFP: membrane fusion protein
MGL : mucus gel layer
mRNA: messenger ribonucleic acid
MS: mass spectrometry
MUC: mucin
NAC: N-acetyl cysteine
NLR: Nod-like receptor
N-WASP: Wiskott-Aldrich syndrome protein
OD: optical density
OM: outer membrane
OML: outer membrane lipoprotein
OMP: outer membrane protein
OMV: outer membrane vesicle
ORF: open reading frame
PAI: Pathogenicity Island
PBS: phosphate buffered saline
PCR: polymerase chain reaction
PE: periplasm
PFA: paraformaldehyde
PMSF: Phenylmethylsulfonyl fluoride
RFU: relative fluorescence unit

RNA: ribonucleic acid
Sat: secreted autotransporter toxin
SCL: subcellular location
SD: standard deviation
SDS-PAGE: sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEM: standard error of the mean
ST: heat-stable enterotoxin
STEC: Shiga-toxin producing *E. coli*
Stx: Shiga toxin
TAM: translocation and assembly module
TIR: translocated initmin receptor
TLR: Toll-like receptor
TMD: transmembrane domain
TNF-α: tumor necrosis factor α
tRNA: transfer ribonucleic acid
TSS: type secretion system
UPEC: uropathogenic *Escherichia coli*
UTI: urinary tract infection
VF: virulence factor

ABSTRACT

Shiga toxin-encoding *Escherichia coli* (STEC) regroup strains that carry genes encoding Shiga toxin (Stx). Among intestinal pathogenic *E. coli*, enterohaemorrhagic *E. coli* (EHEC) constitute the major subgroup of virulent STEC. EHEC colonizes the large intestine and causes diarrhea, hemorrhagic colitis and in some cases, life-threatening hemolytic-uremic syndrome (HUS). The lack of an effective clinical treatment, sequelae after infection and mortality rate in humans, allied with high prevalence of antibiotic resistance make this pathogen as priority for the world health agencies. Initial adherence to host cells is the crucial and first step of the infection of enterohaemorrhagic *Escherichia coli* (EHEC). The attaching/effacing lesions are considered to be important in EHEC. However, although EHEC produce this lesion on animal mondels and cultured human cells, this phenotype has not been demonstrated on human intestinal mucosal surfaces. The aim of the present research work was to identify some novel adhesins in EHEC O157:H7 in a strategy for preventing intestinal colonisation. Rather than classical reverse vaccinology approach solely based on genomic analysis aiming at identifying potential antigen candidates based on the prediction of their subcellular location at the bacterial cell surface, the identification of candidates was here primarily based on subproteomic analysis. To investigate the cell surface proteome, an original approach based on the use of biotinylation reagents of different sizes was first developed. This method highlighted that protein labelling using biotinylation reagents of different sizes provides a sophisticated and accurate way to differentially explore the cell envelope proteome of lipopolysaccharidic diderm bacteria. Following the secrectome concept, three subproteomes were considered and characterised by relative quantification analysis, namely, the proteosurfaceome (proteome at the cell surface), exoproteome (proteome in the extracellular milieu) and proteovesiculome (proteome of the membrane vesicles). This secretome analysis revealed unique proteins expression depending on growth conditions and related to different secretion pathways from EHEC O157:H7, particularly Type III secretion system. Proteomics analysis was combined with functional protein characterisation to identify a group of proteins potentially involved in EHEC O157:H7 adhesion to extracellular matrix and to intestinal cell line. A group of outer membrane proteins revealed important role in the adherent phenotype of EHEC O157:H7 as well promoting bacterial aggregation. Moreover antibodies raised against selected recombinant proteins showed to effectively reduce the adhesion of EHEC to an intestinal cell model.

RIASSUNTO

Shiga toxin-encoding *Escherichia coli* (STEC) contiene ceppi che portano geni codificanti per la Shiga toxin (Stx). Tra i ceppi di *E. coli* patogeni per l'intestino, *E. Coli* enteroemorragico (EHEC) costituisce il gruppo più grande di STEC virulenti. EHEC colonizza l'intestino crasso e causa diarrea, colite emorragica e in alcuni casi, sindrome emolitico uremica. La mancanza di un trattamento clinico efficace, della presenza di complicazioni dopo l'infezione e dell'alto tasso di mortalità negli umani, associato con l'elevata prevalenza di antibiotico resistenza fa di questo patogeno una delle priorità per l'organizzazione mondiale della sanità (OMS). L'adesione alle cellule ospiti è la fase cruciale per l'infezione di *Escherichia coli* enteroemorragico (EHEC). Le lesioni attaching/affacing, sono considerate importanti per EHEC. Comunque, sebbene EHEC produce queste lesioni nei modelli animali e sulle cellule umane coltivate, questo fenotipo non è stato dimostrato sulla superficie delle mucose intestinali. Lo scopo del presente lavoro di ricerca è di identificare nuove adesine in EHEC 0157:H7 per prevenire la colonizzazione intestinale. Piuttosto che sul classico approccio di reverse vaccinology, impostato solamente sull'analisi genomica per identificare potenziali antigeni candidati in base alla predetta localizzazione subcellulare sulla superficie batterica, l'identificazione dei candidati era basata inizialmente sull'analisi subproteomica. Per investigare il proteoma della superficie cellulare, è stato sviluppato un approccio originale basato sull'uso di reagenti biotinilati di diversa grandezza. Questo metodo evidenzia che la proteina marcata usando reagenti biotinilati di diversa grandezza fornisce un sofisticato e accurato modo di esplorare il proteoma della superficie dei batteri gram negativi. Seguendo il concetto di secretoma, tre sotto-proteomi sono stati considerati e caratterizzati con un'analisi quantitativa e relativa, chiamata, proteosurfaceome (proteoma della superficie cellulare), esoproteoma (proteoma nell'ambiente extracellulare) e proteovesiculome (proteoma delle vescicole di membrana). Quest'analisi del secretoma ha rivelato l'espressione di proteine dipendente dalle condizioni di crescita e al differente meccanismo di secrezione in EHEC 0157:H7, in particolare il sistema di secrezione di tipo III. Analisi proteomiche sono state combinate con la caratterizzazione funzionale di proteine per identificare un gruppo di proteine potenzialmente implicate nell'adesione di EHEC 0157:H7 alla matrice extracellulare o alle linee cellulari intestinali. Un gruppo di proteine della membrana esterna hanno rivelato avere un ruolo importante nel fenotipo adesivo di EHEC 0157:H7 promuovendo anche l'aggregazione batterica. Inoltre, anticorpi elicitati contro proteine ricombinanti selezionate hanno mostrato ridurre in modo efficace l'adesione di EHEC ad un modello cellulare intestinale.

CHAPTER I: INTRODUCTION

1. INTRODUCTION

Escherichia coli are inhabitant of the intestines of warm-blooded animals, including humans [1]. *E. coli* typically coexists with the host as a commensal with mutual benefit for both organisms [2] and is usually shed in faeces [3]. Respective to its prevalence, it rarely causes disease but, based on clinical syndromes, the pathogenic strains can be broadly discriminated into intestinal pathogenic *E. coli* (InPEC) resulting in diarrhea, and extraintestinal pathogenic *E. coli* (ExPEC) responsible for urinary tract infections, sepsis and/or meningitis [4]. In humans, the greatest disease burden from *E. coli* results from intestinal infections with severity ranging from mild watery diarrhea and abdominal cramps that proceed to haemorrhagic colitis and severe complications that may lead to death [5]. InPEC represents a leading cause for traveler's diarrhea but also pediatric disease in developing countries, and an emerging source of diarrhea in industrialised countries [6]. While it mainly results from consumption of food-products exposed to direct and/or indirect fecal contamination, six diarrheagenic *E. coli* pathotypes are currently recognised and based on a combination of clinical, epidemiological and molecular criteria [3, 4]. This concept of different *E. coli* enteropathotypes has been originally developed by Nataro and Kaper [3, 4]. *E. coli* possess a core genome, containing the essential genes necessary for growth and survival. The genome contains multiple insertions sites for foreign DNA, where bacteriophages, genomic islands and pathogenicity islands can integrate and define new phenotypic characteristics for the host [7, 8]. Moreover, plasmids contribute to the horizontal gene transfer and the creation of novel combinations of virulence factors leading to new phenotypes and thus the emergence of new pathogens [3, 9]. Besides, some of these virulence factors are horizontally transferred due to the promiscuitive nature of *E. coli*. During evolution, numerous gene exchange events have occurred, resulting in the appearance of new pathotypes that cause new types of disease and outbreaks [10, 11]. Enterotoxigenic *E. coli* (ETEC), enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) are the most prominent *E. coli* enteropathotypes that occur over recent years [12, 13].

1.1. Shiga toxin-producing *E. coli* and enterohaemorrhagic *E. coli*

Shiga toxin-producing *E. coli* or more accurately Shiga toxin-encoding *E. coli* (STEC) comprise a group of bacteria that carry Shiga toxin genes and/or have been shown to produce one or more

Shiga toxins [4]. The term Shiga toxin has been originally introduced by O'Brien and colleagues for *E. coli* due to the molecular and functional similarity to Shiga toxin (Stx) encoded by the genetically closely related *Shigella dysenterieae* type 1 [14]. Of note, converging evidences, such as the most recent whole genome sequence analyses, clearly indicates now that *Shigella* are actually *E. coli* [15] and are related to EIEC causing varying degrees of dysentery [16]. A large amount of different *stx* subtypes with different designations has been detected in *E. coli* during the last decades, and in a recent large multicenter study, a unified Stx nomenclature has been recommended [17]. This approach aimed at simplifying the nomenclature of these toxin groups but some scientists also use old-style designations such as verotoxins or verocytotoxins [18, 19].

Several subgroups of STEC are defined and EHEC is the most important group. From early definitions, EHEC are pathogenic STEC that are able to cause the typical symptoms from watery diarrhea, bloody diarrhea, up to the haemorrhagic colitis (HC), the haemolytic uremic syndrome (HUS), and/or other extraintestinal sequelae in humans [20]. Little progress has been made in reducing the huge number of infections associated with these pathogens and few interventions to reduce the food contamination and the infectious complications of this enteric disease are available [21]. While the vast majority of EHEC infections are sporadic, they can also lead to major outbreaks worldwide [22]. Of note, the ruminants are the natural reservoir of EHEC but they are asymptomatic carriers. EHEC are characterised by Stx-production and the formation of attaching and effacing (A/E) lesions encoded by a 35-kb chromosomal pathogenicity island (PAI) known as the LEE (locus of enterocyte effacement), as demonstrated in human cultured cells [4, 23, 24]. There are over 380 distinct serotypes of STEC [25, 26] but only EHEC of serotypes O157:H7, as well as the big six serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19 and O145:H28 are the most frequently associated with human disease [27]. LEE-positive EHEC and LEE-positive EPEC are sometimes referred to A/E encoding *E. coli* (AEEC) [28]. Interestingly, EHEC O157:H7 would have evolved from EPEC O55:H7 following a series of stepwise acquisitions and losses of genes [29]. The terms AEEC and STEC rely mainly on the genetic traits of the bacteria (genotype), *i.e.* harboring *eae* (enterocyte *attaching* and *effacing*) or *stx* respectively, whereas EHEC and EPEC are primarily classified on the clinical symptoms of the patients (phenotype) and in a second line on the molecular traits [30]. Recently, LEE-negative STEC and EHEC strains have also been reported [31]. LEE-negative STEC occur mainly in animals or the environment, and are frequently isolated from food sources. LEE-negative EHEC are more rarely isolated from patients with HUS. However, a LEE-negative *E. coli* O104:H4 has caused the large outbreak in Europe in 2011 and appeared as a hybrid of EAEC and EHEC [32]. These *E. coli* O104:H4 strains have been tentatively named

enteroaggregative haemorrhagic *E. coli* (EAHEC) [33] and can be categorised as a subgroup of EHEC.

1.2. Enterohaemorrhagic *E. coli*: intestinal colonisation and host response

Enterohaemorrhagic *E. coli* is well distributed in diverse environments such as water, soil, plants and animal reservoirs which play essential role in its persistence and dispersal [34]. It is a commensal bacterium commonly present in the gastrointestinal tract of ruminants such as cattle, goats, sheep, and deer [35]. Typically, this zoonotic agent spreads to human through the ingestion of undercooked meat, ingestion of ruminant feces-contaminated foodstuffs such as fruits, vegetables, drinking contaminated water, person-to-person contact, and hospital-acquired nosocomial infections [36]. Until colonise the follicle-associated epithelium of Peyer's patches and villi of the terminal ileum and colon, EHEC have to overcome chemical, mechanical and biological barriers [37]. First, it has to survive to the chemical host protection such saliva containing mucins and enzymes, acid stress in the stomach, bile secretion in the small intestine, and antimicrobial peptides throughout the intestine. Once in intestine EHEC encounters a mechanical barrier composed by the mucus layer and the intestinal microflora, that associated with the innate and acquired immune response to the pathogen represent host first line of defense [38]. The interaction between microorganism and host cells comprises multiple stages, classical divided in three steps. EHEC pathogenesis involves the initial adherence of the bacterium to the host's intestinal epithelium. At this stage, EHEC form microcolonies on the surface of tissue culture cells in a pattern known as localised adherence [39]. EHEC come in contact with short-chain fatty acids, acetate, propionate, and butyrate, secreted by the intestinal microbiome as fermentation products. Butyrate and high concentrations of short-chain fatty acids trigger expression of the bacterial genes such *lrp* and *iha* conferring colonic adherence, motility and formation of attaching and effacing (A/E) lesions [40]. The interaction with the intestinal commensal microflora also activates communication between bacteria known as quorum sensing and between bacteria and host hormones. Host catecholamine promotes virulence by activating bacterial motility (flagellar synthesis), formation of A/E lesions, and increased expression of Stx [41]. The second stage of EHEC pathogenesis is mediated mainly by a Type III secretion system (T3SS), which translocates bacterial proteins including EspA, EspB, and EspD into the epithelial cells. Like most of the structural proteins of the systems, these effectors are encoded within the LEE pathogenicity island by the *esc* and *sep* genes [42]. The complete sequence of LEE region has demonstrated the

codification of a T3SS, as well as other genes necessary for pedestal formation such T3SS-secreted proteins, named Esps (EPEC-secreted proteins), including EspA, EspB, EspD, and EspF, as well as an adhesin, intimin, and its translocated receptor, Tir. Mutation of any of these bacterial factors, with the exception of EspF, prevents A/E lesion formation in epithelial cell culture models [4, 42, 43].

The third step of EHEC infection is characterised by enterocyte effacement, pedestal formation, and intimate bacterial attachment to the host cell. Briefly this step requires two essential proteins: the outer bacterial membrane protein, intimin and its receptor Tir [44]. Once secreted into the host cytoplasm, Tir is placed on the host cytoplasmic membrane as a hairpin-like structure, with its N- and C-terminus in the cytoplasm and central domain exposed to the surface which interacts with intimin to form a tight attachment of the bacteria to the eukaryotic cell [44]. In cooperation with non-LEE encoded F-like protein from prophage U (EspFu) recruits host proteins such actin nucleation-promoting factor Wiskott-Aldrich syndrome protein (N-WASP) and insulin receptor tyrosine kinase substrate p53 (IRSp53) to subvert host cytoskeleton and actin polymerisation resulting in actin accumulation and consequently formation of pedestal-like structure (Figure 1) [45, 46].

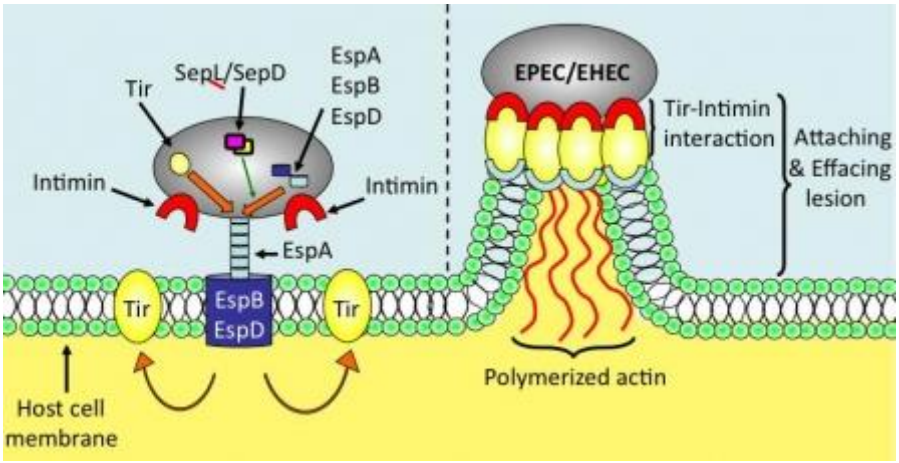


Figure 1. Schematic representation of the formation of attaching and effacing (A/E) lesions. Bacterial effectors are secreted into the host cell. Tir directs to the cell surface promoting actin polymerisation. (Adapted from Shantanu Bhatt, Ph.D. Saint Joseph's University).

As introduced above, during intestinal colonisation the host naturally possesses mechanisms to resists to infection. The intestinal inflammatory response is a paramount feature of host resistance to infection. The activation of innate immune response uses pattern recognition receptors to detect pathogen-associated molecular patterns expressed by bacteria to trigger an effective protective antimicrobial immune response [34]. Toll-like receptors (TLRs) and Nod-like receptors (NLRs) are the principal pathogen recognition receptors involved in recognizing and responding to bacterial antigens. Particularly, TLRs are a family of membrane-bound receptors that recognise specific microbial components such as Gram-negative bacterial lipopolysaccharide (TLR 4) and flagellin (TLR 5) [47]. General activation of the immune system is triggered by the activation of TLRs and NLRs, this involves the release of immune mediators and immune cell activation that collectively serves to contain and clear the bacterial infection such the interferon-gamma (IFNγ), through Jak 1, 2, Stat1 signal transduction pathways [48]. Macrophages, natural killer T cells, and activated T cells secrete cytokines, including IFNγ into the extracellular environment after bacterial infection resulting in the activation of IFNγ-stimulated genes in recipient host cells that together make the host defense against EHEC [49]. Indeed the role of IFNγ signal transduction pathway is critical as response to bacterial infections. Severe recurrent life-threatening microbial infections results of the defects in any part of the IFNγ signaling cascade [50]. This important role in immune response makes IFNγ pathway an ideal target for bacterial subversion; for example, in *Listeria monocytogenes*, the upregulation of suppressors of cytokine signalling activity results in downregulation of IFNγ activation [51]. EHEC also seems to suppress the IFNγ signal transduction pathway via their Shiga toxin [52].

1.3. Colonisation factors of STEC from a secretome perspective

Basically, pathogenic organisms differentiate from their nonpathogenic counterparts by encoding specific virulence factors secreted to the bacterial cell surface or beyond [53]. To date, nine protein secretion systems, numbered from Type I to Type IX secretion system (T1SS to T9SS), have been uncovered in Gram-negative bacteria [54]. Considering the wealth of systems and effectors, the secretome is a powerful concept to apprehend protein trafficking as a whole in living cells by considering both the secretion routes and their cognate secreted proteins [55, 56]. While the LEE is a key and prominent molecular determinant in the pathogenicity, neither all EHEC nor STEC possess the LEE, suggesting that they possess additional virulence and colonisation factors [57]. The colonisation processes are essential in human infection but also for bacterial multiplication in

animals and contamination of foodstuffs. In order to highlight the molecular mechanisms responsible for their correct expression and subcellular localisation, those effectors are stated here from a protein secretion system perspective.

Protein secretion is an essential cellular function present in all living cells. In bacteria, it is involved in a huge diversity of functions playing important roles in the bacterial physiology from adaptation and response to environmental cues, adhesion and biofilm formation, to pathogenicity and virulence [54]. While secretion refers to the active transport from the interior to the exterior of the cell compartment, secreted proteins have three main possible fates in diderm-LPS bacteria (archetypal Gram-negative bacteria): (i) remain associated with the bacterial outer membrane (OM), (ii) be released into the extracellular milieu, or (iii) be injected into a target cell (either a eukaryotic or bacterial cell) [58]. Among the nine protein secretion systems in diderm-LPS bacteria secreted proteins [54], T1SS to T8SS can be involved in surface colonisation of STEC (Figure 2). Respective to those systems, the secreted proteins involved in the surface colonisation process, namely adhesion and biofilm formation, are reviewed in STEC (Table 1). In addition, some outer membrane proteins (Omp) assembled by the Bam (β-barrel assembly machinery) complex [59], namely IhA (IrgA homologue adhesion) [40, 60] or OmpA [61], can also participates to the colonisation process in STEC.

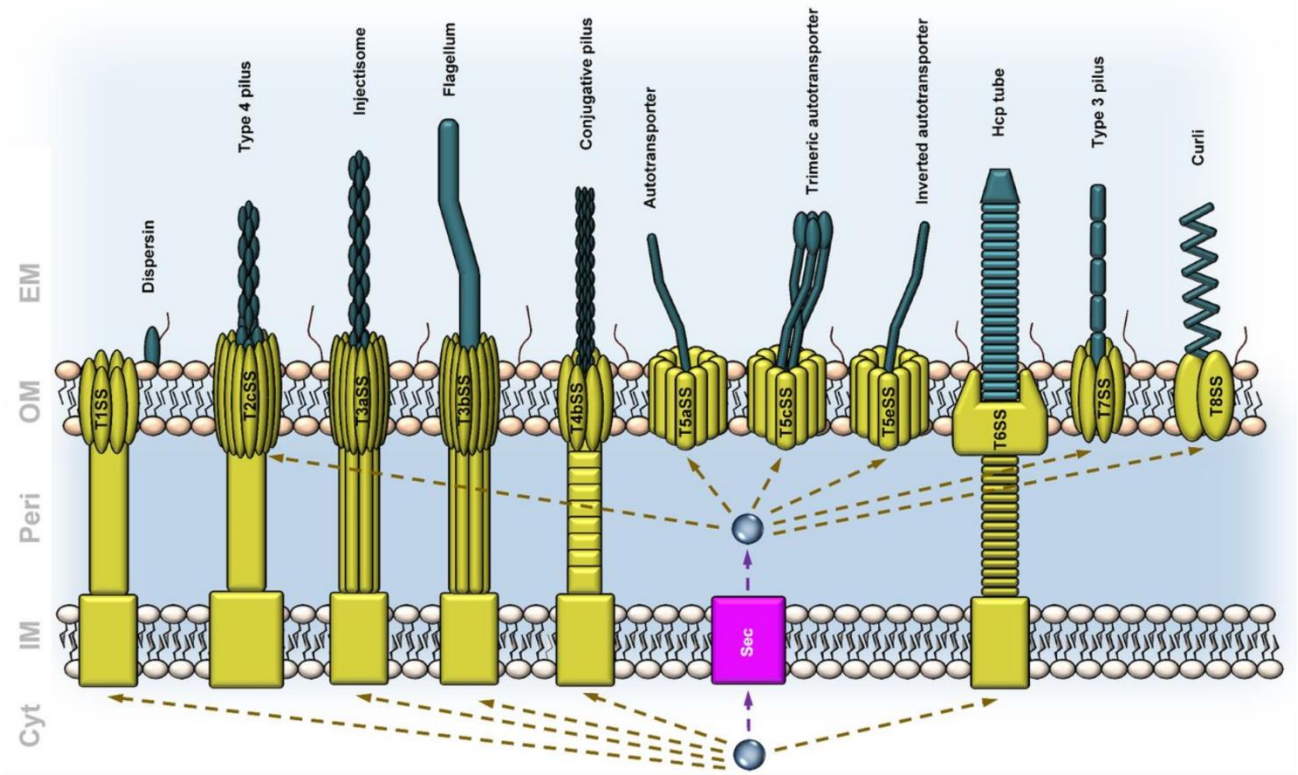


Figure 2. Overview of the complement of the secretome associated with the surface colonisation factors in STEC. Protein export systems are coloured in violet, protein secretion systems are coloured in green and their cognate effectors directly involved in surface colonisation are coloured in blue. For details of the respective secretion system and cognate colonisation factors, refer to the main text. T1SS: type I protein secretion system or the ABC-MFP-TolC heterotrimeric system. T2SS: type II protein secretion system or the SDP, especially the T4P system (T2cSS). T3SS: type III protein secretion system, including the injectisome (T3aSS) and flagellar (T3bSS) systems. T4SS: type IV protein secretion system, especially the conjugative Tra system (T4bSS). T5SS: type V protein secretion system, especially including the autotransporter (T5aSS), trimeric autotransporter (T5cSS) and inverted autotransporter (T5eSS) systems. T6SS: type VI protein secretion system. T7SS: type VII protein secretion system or the CU pathway. T8SS: type VIII protein secretion system or the ENP pathway. Cyt, cytoplasm; IM, inner membrane; Peri, periplasm; OM, outer membrane; EM, extracellular milieu [62].

Table 1: Colonisation factors of STEC, including EHEC and related enteropathotypes, according protein secretion systems in diderm-LPS bacteria [62].

T1SS [a]	**T2SS** [b]	**T3SS** [c]		**T4SS**	**T5SS** [d]			**T6SS** [e]	**T7SS** [f]	**T8SS**	**Non-classical secretion** [g]
	T2cSS	**T3aSS**	**T3bSS**	**T4bSS**	**T5aSS**	**T5cSS**	**T5eSS**				
Aap	BFP	Injectisome	Flagellum	Conjugative	EhaA	EhaG	Intimin	Hcp tube	ECP	Curli	GAPDH
	HCP	Tir		pilus	EhaB	Saa	FdeC		Lpf		
					EhaD	EibG			ELF		
					EhaJ				SFP		
					Cah				F9		
					EspP				AAF		
					EspC						
					PssA						
					Sab						

[a] Aap: dispersin, anti-aggregation protein.

[b] BFP: bundle-forming pili; HCP: haemorrhagic coli pili.

[c] Tir: translocated intimin receptor.

[d] Eha: enterohaemorrhagic *E. coli* autotransporters; Cah: calcium-binding antigen 43 homologue; EspP: extracellular serine protease plasmid-encoded; EspC: EPEC-secreted proteins C; PssA: the protease secreted by STEC; Sab: STEC autotransporter. Saa: The STEC autoagglutinating adhesion autotransporter; EibG: *E. coli* immunoglobulin-binding protein G; FdeC: factor adherence *E. coli*.

[e] Hcp: haemolysin-coregulated protein.

[f] ECP: *E. coli* common pilus; Lpf: Long polar fimbria; ELF: *E. coli* YcbQ laminin-binding fimbria; SFP: sorbitol-fermenting fimbriae; AAF: aggregative adhesion fimbria.

[g] Non-classical secretion refers to unkown secretion system (*i.e.* as yet uncovered) but, in the end, it could also correspond to protein trafficking mechanisms unrelated to protein secretion; GAPDH: glyceraldehyde 3-phosphate dehydrogenase.

1.3.1. Type I secretion system (T1SS): Dispersin

The T1SS refers to a heterotrimeric complex constituted by an inner membrane ABC exporter, a membrane fusion protein (MFP) and a TolC-like pore-forming Omp [63]. The T1SS secretes proteins in a single step directly from the cytosol to the extracellular milieu, without periplasmic intermediates. In some EAEC, a T1SS is encoded in the enteroaggregative ABC transporter locus (*aat*) present in a large virulence plasmid called pAA (aggregative adherence) [64]. This system is involved in the secretion of the Aap (anti-aggregation protein) dispersin at the cell surface following non-covalent binding to lipopolysaccharides (LPS) [65]. By preventing hyper-aggregation of bacteria and collapse of aggregative adherence fimbriae (AAF), dispersin allows bacterial cells dispersion along the host intestinal mucosa, thus contributing to the adherence and colonisation of EAEC [66]. As it is found in some commensal *E. coli* but is also absent from some EAEC [67], the role of dispersin in pathogenesis is not clear. Nonetheless, dispersin Aap is present in some STEC, such as EHEC O104:H4 [68].

1.3.2. Type II subtype c secretion systems (T2cSS): Type 4 pili

The T2SS, or secreton-dependent pathway (SDP), is responsible for the secretion of proteins first exported via the Sec or Tat systems [69]. While the T2SS is primarily involved in the secretion of free soluble extracellular proteins, namely through the subfamilies Xcp, *i.e.* the T2SS of subtype a (T2aSS), and Hxc, *i.e.* the T2SS of subtype b (T2bSS), it can also be involved in pili assembly [70]. Indeed, the system responsible for the subunit secretion, assembly and biogenesis of the type 4 pili (T4P) is a distant homologue to the paradigm Xcp T2aSS, and actually corresponds to the Type II subtype c secretion system (T2cSS) [54, 71]. In pathogenic *E. coli*, T4P are involved in bacterial virulence, including the colonisation process, especially host cell adherence, biofilm formation, bacterial aggregation and the twitching motility [72]. In EPEC, the bundle-forming pili (BFP) are responsible for initial bacterial attachment to the intestinal epithelium and formation of compact microcolonies [73-75]. BPF are encoded on the plasmid pEAF (EPEC adherence factor) and are therefore not systematically present in all EPEC [76, 77]. In STEC, the T4P named HCP (haemorrhagic coli pili) is composed of a 19-kDa pilin subunit encoded by the *hcpA* chromosomal gene, also called *ppdD* (prepilin peptidase-dependent) in *E. coli* K-12. In EHEC O157:H7 EDL933, HcpA is directly involved in cell adherence and in the ability to invade human and bovine epithelial cells. HCP also contributes to the biofilm formation due to its ability to bind some extracellular matrix (ECM) proteins, especially fibronectin and laminin [78]. In some EAEC, T4P are encoded

on conjugative plasmids (thus encoding a T4bSS in addition, see below) and contribute to bacterial adherence to epithelial cells and abiotic surfaces, as well as biofilm formation [79].

1.3.3. Type III secretion systems (T3SS): Injectisome and flagella

The injectisome and the flagellum are assembled by two phylogenetically distinct but homologous Type III secretion systems (T3SS) of subtypes *a* and *b*, i.e. T3aSS (injectisome system) and the T3bSS (flagellar system) respectively [80, 81]. Among InPEC, the T3aSS is the hallmark of EPEC but it is also present in some but not all EHEC [82]. The T3aSS consists of a double-membrane embedded nanomachines and promotes the delivery of bacterial effectors to the cytoplasm or plasma membrane of target eukaryotic cells, where they can modulate or subvert a large variety of host cell mechanisms but also promote bacterial invasion and colonisation [83]. The injectisome formed by the T3aSS is a needle structure, a multiring complex that spans the bacterial envelope and plasma membrane of the host cell providing a continuous path up to the cytosol of the infected host cell. In EPEC, the T3aSS is directly involved in cell adhesion and pedestal formation resulting in characteristic histopathological A/E lesions [84]. Contrary to what is sometimes wrongly assumed by part of the scientific literature, such lesions are never observed from clinical samples after EHEC infection [4, 23, 24]. Nonetheless, A/E lesions are frequently observed *in vitro* from EHEC infected epithelial cell cultures. The T3aSS structure is encoded by genes located in the LEE pathogenicity island [85]. Among all the proteins secreted by the injectisome T3aSS, Tir (translocated intimin receptor) is the primary molecule associated with intimate bacterial interaction with the epithelia and A/E lesions [86]. Tir, encoded by the *espE* gene located in the LEE, is secreted and injected into the host cell by the injectisome T3aSS, and later localised at the plasma membrane of the infected host epithelial cell. This bacterial protein exposed at the host cell surface then acts as a receptor for a direct and specific binding of the bacterial cell-surface exposed intimin. By protein-protein interaction, Tir activates the recruitment and rearrangement of the host-cell cytoskeletal actin, which rearranges and results in the formation of a characteristic pedestal structure [87]. Furthermore, the injectisome plays a role in the adhesion to plants with a marked tropism for the stomata [88, 89]. Besides adhesion to the host cell, the T3aSS also participates to invasion capabilities in EIEC [77, 90]. While T3aSS is only present in a subset of InPEC, namely EPEC, EIEC and some EHEC, flagellum is quite ubiquituous across *E. coli*. In this species, the flagella are peritricheous forming a ponytail when in motion [91]. The different components of the flagellum are secreted by the T3bSS, namely the hook-filament junction protein, the filament-

capping protein, and flagellin, the major subunit of the filament [92, 93]. Of course, flagella are primarily involved in cell motility, especially swimming and/or swarming in *E. coli* depending on environmental conditions [94]. Besides coordinated movement at surfaces contributing to colonisation, that is swarming, flagella participate to adhesion and invasion by providing motility towards surfaces or target cells [95]. In EHEC, flagella play a role in the adhesion to the epithelial cells but are down regulated after contact with the epithelium and would then just initiate early stages of the adhesion process [96]. In EPEC, the flagellum tip protein FliD can mediate adhesion to Caco-2 cells. Following vaccination targeting the flagellin FliC, the colonisation of EHEC in cattle was significantly decreased providing evidence of the importance of the flagella in host gut colonisation [97]. Most EIEC, though, have no flagellum and, consequently, are nonmotile with no H antigen serotype [77].

1.3.4. Type IV subtype b secretion system (T4bSS): Conjugative pili

Respective to protein secretion, the T4SS can be discriminated into two subtypes. The T4aSS includes effector translocator systems homologous to the prototypical VirB/D4 complex extensively investigated in *Agrobacterium tumefaciens*, notably involved in the injection of effectors into infected host cell. The T4bSS comprises the conjugation machines homologous to the prototypical F-episome conjugal transfer (Tra) system, notably involved the transport of nucleoprotein complexes [98-100]. While no T4aSS has been described so far in *E. coli*, conjugative plasmids can be present [101]. Besides pO157 [102], plasmid profiling in different *E. coli* O157:H7 strains indeed revealed the presence of numerous additional plasmids, which highly vary in size and number [103, 104]. Among them, pO157-Sal was identified and demonstrated to be a novel conjugative plasmid of the Incl family in STEC [105, 106]. This plasmid contains the full set of *tra* genes and thus encodes a T4bSS. From what is known of the F episome in *E. coli* K-12, such a plasmid could further contribute to bacterial colonisation [107]. Indeed, conjugative F pili assembled by the T4bSS would induce biofilm formation by improving adhesion capability of the bacterial cells. Horizontal transfer of the F episome to siblings within the biofilm further increase the proportion of transconjugant cells and further expands the colonisation propensity of the bacterial population [107-109].

1.3.5. Type V secretion systems (T5SS): Adhesins by numbers

The T5SS can be categorised into 5 subtypes, (i) the autotransporter system (subtype a), (ii) the two-partner secretion pathway (subtype b), (iii) the trimeric autotransporter system (subtype c), (iv) the hybrid autotransporter system (subtype d), and (v) the inverted autotransporter system (subtype e). Broadly speaking, the T5SS refers to protein secretion *via* an OM pore formed by a β-barrel, the secreted proteins being first exported *via* the Sec translocon. Except for the T5dSS, all four other subtypes can secrete proteins involved in surface colonisation. Although a T5bSS is present in *E. coli* 0157:H7 [110], the exoprotein OtpA (O157:H7 two-partner protein A) does not display any adhesive properties towards intestinal epithelial cells and the possible contribution of this secretion system and cognate secreted proteins to surface colonisation remains to be established.

1.3.6. Type V subtype a secretion system (T5aSS): SAAT and SPATE adhesins

Autotransporters are single polypeptides that can drive their own secretion through the OM *via* a C-terminal translocator forming a β-barrel in the OM allowing translocation of the N-terminal passenger domain. The main function of T5SS is to secrete virulence factors but it also participates in cell-to-cell adhesion [111]. The self-associating autotransporters (SAATs) are cell-surface exposed outer-membrane proteins (OMPs) encompassing AIDA (adhesin diffuse adherence), TibA (enterotoxigenic invasion locus b protein A) and Ag43 (antigen 43) [112, 113] autotransporter family. Besides autoaggregation, SAATs are actively involved in bacteria-host interaction and biofilm formation [114]. Five autotransporters belonging to the AIDA family have been identified in *E. coli* 0157:H7 and are called enterohaemorrhagic *E. coli* autotransporters (Eha) [115]. The EhaA overexpression in *E. coli* K-12 appeared to promote autoaggregation, biofilm formation and adhesion to bovine primary epithelial cells from the terminal rectum [115]. EhaB specifically binds to some ECM proteins, especially laminin and collagen I [116]. Similarly, EhaJ mediates specific adhesion to collagens I, II, III and V as well as fibronectin, fibrinogen and laminin [117]. While EhaD and EhaJ also promote biofilm formation, the possible function of EhaC in the colonisation processes remains to be established [115]. Interestingly enough, glycosylation of EhaJ appeared of importance for biofilm formation but not for binding to the ECM [117]. While no TibA homolog has been reported in STEC, *E. coli* O157:H7 encodes an Ag43 homolog called Cah (calcium-binding antigen 43 homologous) promoting cell autoaggregation but apparently not involved in adhesion to the host cells [118]. The SPATEs (serine protease autotransporters of *Enterobacteriaceae*) constitutes another autotransporter family generally released in the

extracellular milieu and primarily with proteolytic activity against various substrates but also exhibiting adhesion properties for some of them [119]. In *E. coli* O157:H7, EspP (extracellular serine protease plasmid-encoded) is directly involved in biofilm formation but also adherence to epithelial cells [120, 121]. Recently, EspP was shown to oligomerise to form megastructural ropes, which possess adhesive and cytopathic activities on host epithelial cells [121, 122]. In EPEC, EspC oligomerises similarly and serves as substratum for bacterial adherence and biofilm formation [122]. In STEC O26, PssA (protease secreted by STEC) was shown to participate to intestinal colonisation of calves [123]. In STEC O113, Sab (STEC autotransporter contributing to biofilm formation) is involved in adherence to abiotic surfaces and epithelial cells [124]. Of note, the high variability in the presence of those different autotransporters in STEC may significantly contribute to difference in colonisation abilities and even modulate virulence. For example, the identification of Sab in LEE-negative STEC O113:H2 contrasts with its absence from a LEE-positive strain collection and would suggest an alternative way to adhere to the host cells for strains defective in their ability to induce A/E lesions [124].

1.3.7. Type V subtype c secretion system (T5cSS): Trimeric autotransporter adhesin EhaG

The T5cSS corresponds to autotransporters formed upon homotrimerisation. All trimeric autotransporters characterised to date are exposed to the bacterial cell surface and play a role in adhesion [125-127]. EhaG (EHEC adhesin G) is present in diarrheagenic *E. coli*, from EHEC, ETEC, EPEC, EAEC to EIEC, and is highly prevalent in STEC [128]. Upon overexpression in *E. coli* K-12, EhaG from *E. coli* O157:H7 was shown to mediate autoaggregation resulting in the formation of strong biofilm. In addition, adhesion to intestinal epithelial cells and specific binding to collagens I, II, III, and V as well as to laminin, fibronectin and fibrinogen was reported [128]. Saa (STEC autoagglutinating adhesin) is involved in adhesion to intestinal epithelial cells and was identified in various LEE-negative STEC [57]. EibG (*E. coli* immunoglobulin-binding protein G) contributes to a chain-like adhesion pattern to human epithelial cells also in LEE-negative STEC [129, 130].

1.3.8. Type V subtype e secretion system (T5eSS): Intimin

The T5eSS refers to inverted autotransporters in the sense that the translocator is located at the N-terminal instead of the C-terminal region of the monomeric autotransporter [131-133]. Intimin from

EPEC and EHEC is a prototypical member of this OMP family. Along with the T3aSS, intimin is encoded by the *eae* gene in the LEE and, as already mentioned above, interacts specifically with Tir resulting in intimate attachment of the bacteria with the host-cell surface, pedestal formation and A/E lesions [7]. While five alleles (α, β, γ, δ and ε) have been reported for *eae* with a total of 27 variants [134, 135], these intimins appeared functionally interchangeable [87, 136]. Besides Tir, the binding of intimin to alternative receptors such as β_1 integrins or nucleolin remains unclear [131, 137]. Nonetheless, intimin can also contribute to intestinal colonisation independently from its ability to bind to Tir [136].

While no invasin has been reported to date in STEC [131], FdeC (factor adherence *E. coli*) was recently uncovered as another representative of the T5eSS and appeared widely distributed in *E. coli*, including InPEC, *e.g.* STEC and EHEC [138, 139]. FdeC binds to human epithelial cell and exhibits affinity for some ECM components, notably collagens I, III, V and VI [138]. Its contribution to the kidney and bladder colonisation was associated to the propensity of UPEC bacterial cells for autoaggregation. In EHEC O26:H11, the expression of FdeC regulated by H-NS (histone-like nucleoid-structuring protein) and occurring at temperature above 39°C was demonstrated to mediate biofilm formation and could contribute to the colonisation of the terminal rectum of cattle [139].

1.3.9. Type VI secretion system (T6SS): Aggregation to host cell surface

The T6SS consists of two principal associated complexes with an additional bridging and cytoplasmic elements: a membrane-associated assembly and an assembly similar to the bacteriophage sheath, tube and spike proteins [140]. In some EHEC, EPEC and EAEC strains, bioinformatic analysis identified more than 10 orthologues of known T6SS components present [141-144], suggesting they could be crucial mediators in the aggregation to the host cell surface [145]. In EAEC, the correct expression of the T6SS apparatus, including the Hcp (haemolysin-coregulated protein) tube forming the inverted phage tail, appeared critical for biofilm formation [146]. Still, the exact contribution of this secretion system in the colonisation process remains to be established in STEC.

1.3.10. Type VII secretion system (T7SS): Pili by numbers

In diderm-LPS bacteria, the T7SS corresponds to the chaperone-usher (CU) pathway involved in the secretion and assembly of pili [147]. The nomenclature for pili formed *via* the T7SS is messy, quite confusing, and essentially species-dependent; *e.g.* under the general denomination of Type 3 pili, the designations of CS (coli surface) pili, P pili, aggregative adherence fimbriae (AafD), adhesive fimbriae on RDEC-1 or diffuse adherence fibrillar adhesin/Dr blood group antigen (Afa/Dr) are also found but all fall under the T7SS umbrella [148-150]. While numerous operons encoding such pili remain putative [151-153], STEC secrete and assemble several pili via the T7SS. The operon encoding ECP (*E. coli* common pilus) is highly prevalent across *E. coli*, including EPEC [154]. These pili participate to bacterial adhesion to human epithelial cells, intestinal colonisation and biofilm development through interorganelle binding via the EcpA pilin [155]. In *E. coli* O157:H7, these pili were further demonstrated to target arabinosyl residues in plant cell walls to mediate adhesion to vegetables [156]. *E. coli* O157:H7 also contains two operons encoding Lpf (long polar fimbriae), namely *lpf1* with six genes (*lpfABCC'DE*) forming LpfA pilus, and *lpf2* containing five genes (*lpfABCDD'*) forming LpfD2 pilus [157]. Regarding the Lpf major subunits, five different genetic variants have been described for LpfA1 against three for LpfA2 [158]. Lpf interact specifically with the ECM, namely fibronectin, laminin or collagen IV components. They contribute to the adhesion of EHEC to the intestinal epithelium and can even influence bacterial tropism towards different intestinal tissues [159, 160]. In EHEC O104:H4, LpfA contributes to bacterial adhesion to polarised intestinal epithelial cells, biofilm formation as well as caecum and large intestine colonisation [161]. In EHEC, ELF (*E. coli* YcbQ laminin-binding fimbria) were found to bind laminin as well as to adhere to human epithelial cells [162]. SFP (sorbitol-fermenting fimbriae protein) are responsible for haemagglutination activity in *E. coli* O157:H7 but also act as adhesin *via* the SfpG pilin, involved in adherence to epithelial cells [163]. F9 pili are encoded in O-island 61 and are involved in binding of EHEC to fibronectin and to bovine intestinal epithelial cells [164]. While encoded in numerous STEC [165], expression of Type 1 pili was only reported in a subset of STEC O26 and O118 strains [166], where they participate to the colonisation of the bovine gut by mediating adhesion to intestinal epithelial cells but also enable adhesion to abiotic surfaces [167, 168]. Actually, the expression of Type 1 pili would be subjected to phase variation of *fimA* encoding the major prepilin but the triggering environmental factors remain to be determined [165, 169].

Like EAEC, EHEC O104:H4 lacks the LEE but encodes typical AAF (aggregative adhesion fimbria) [32, 33]. Just like all EPEC are AEEC, all EAEC are aggregative-adherence encoding *E. coli* (AAEC), *i.e.* aaf^{+}, from which atypical EHEC O104:H4 would have emerged [32, 33, 170]. In EAEC, AAF are associated with a strong ability to form biofilms on biotic and abiotic surfaces as well as haemagglutination with human erythrocyte [43]. Four variants of AAF have been identified, namely AAF/I to AAF/IV, encoded on virulence plasmids of the pAA family [171]. AF/II was further demonstrated to bind fibronectin [172]. AAF act in concert with dispersin in the colonisation of the intestinal mucosa, which would result in a highly virulent combination in EHEC O104:H4 when Stx is also present [173]. While the expression of those different pili is subjected to regulation by various environmental factors [174], their global expression and respective contribution to the colonisation process along the food chain, from the environment, animal reservoirs, food matrices to human infection is far from being understood. Besides, several operons encoding putative T7SS remain to be characterised in STEC.

1.3.11. Type VIII secretion system (T8SS): Curli

The T8SS corresponds to the extracellular nucleation-precipitation (ENP) pathway involved in the secretion and assembly of peculiar pili, called curli [147]. In fact, curli are functional amyloid fibers predominantly composed of the major curli subunit protein CsgA following nucleation at the cell surface initiated by the minor curli subunit CsgB, which further promote ramification along the fibres [175]. Curli fibres are extremely adherent and involved in cell aggregation, bacterial adhesion and, ultimately, biofilm development [176, 177]. *E. coli* O157:H7 expressing curli are more virulent and exhibit higher adherence abilities to eukaryotic cells than noncurliated strains [178]. The expression of curli seems to compensate the absence of Lpf [179]. It also appeared that rather than affecting initial attachment, curli enhanced sessile development [180].

1.3.12. Non-classical secretion and other protein trafficking mechanisms: Glycolytic enzymes

Besides secretion *via* known protein secretion systems, some proteins devoid of any identifiable signal sequence can be identified outside the cytoplasm. In some conditions, it can be reasonably thought those proteins could be secreted by alternative but unknown protein secretion machineries, the so-called non-classical secretion. Besides piggybacking, where a protein like a chaperone

associates with a secreted protein and is collaterally transported, the true secretion mechanism is sometimes uncovered years after, such as for the secreted proteins *via* the most recently uncovered T6SS or T9SS [54]. In parallel, mechanisms unrelated to secretion also participate to protein trafficking, such as (i) cell lysis, which can be controlled (allolysis), results from bacteriophage infection or follows the entry into a lytic cycle (resulting in the extracellular release of Stx), or (ii) membrane budding resulting in OMVs (outer membrane vesicles) released from the bacterial cell surface.

Among unexpected extracytoplasmic proteins, glycolytic enzymes are frequently uncovered [181]. Because they exhibit completely different function when outside the cytoplasm, these proteins are qualified of moonlighting to highlight their hidden second function. In EHEC and EPEC, the GAPDH (glyceraldehyde 3-phosphate dehydrogenase) present at cell surface binds plasminogen and fibrinogen and remains associated with intestinal epithelial cells upon infection [182]. Depending on the growth conditions, secretion of GAPDH could occur *via* piggybacking through the T3aSS [183]. The exact contribution of GAPDH and other putative moonlighting glycolytic enzymes to colonisation in STEC would deserve further investigations.

1.4. Vaccines against pathogenic *E. coli*

The prevention of *E. coli* infections is of pressing concern from both the public health and economic perspectives [184]. Indeed, the high range of diseases caused by *E. coli*, associated with high costs to healthcare systems makes the absence of a broadly protective vaccine against pathogenic *E. coli* strains a major problem for modern society. The overall problem was exacerbated when, in April 2014, the World Health Organisation published the first global report on antibiotic resistance revealing serious, worldwide threat to public health [185]. However, attempts to develop a broadly protective and safe vaccine against *E. coli* have not been successful so far, being the genetic and antigenic variability of pathogenic *E. coli* species the principal obstacle. This variability coupled with the difficulty prediction of vaccine coverage and protection, has led to denying numerous promising pre-clinical candidates by human trials [186, 187]. Also *E. coli* pathotypes have been considered and studied in isolation by the scientific community instead of addressing *E. coli* as a single microorganism responsible for human and animal diseases.

Many efforts have already been done to assess protein moieties as putative vaccine candidates focusing, essentially on surface-exposed proteins potentially involved in pathogenesis, such as adhesins, iron-regulated outer membrane proteins (OMPs) and toxins. Several iron-regulated OMPs

have also been assessed to date as potential vaccine candidates, given that many are surface-exposed and iron acquisition is a requisite for pathogenesis [188]. The pore-forming toxin α-hemolysin has been demonstrated to be highly conserved [189]. In a mouse model of pyelonephritis, systemic immunisation with purified HlyA was associated with decreased renal damage, but did not affect clearance of *E. coli*. However, its combination with digalactoside-binding pilus was able to prevent both bacterial colonisation and renal injury [188, 190].

An advantage on targeting adhesins is the promise to enhance the bactericidal activity mediated by complement and phagocytes as well to inhibit bacterial colonisation of host structures, a critical step in *E. coli* infection [188]. For example, systemic immunisation with purified P fimbriae [190, 191] and synthetic peptides corresponding to the protective epitope of the P fimbrial major subunit PapG conferred protection in a murine pyelonephritis model [192]; immunisation with purified P fimbriae or with purified PapDG-complex also conferred protection in a nonhuman primate model [193, 194].

Another approach to prevent extraintestinal *E. coli* uses whole-cells for immunisation. This method has high potential, since it could present multiple antigens, elicit antibodies against conformational and linear epitopes and possess natural adjuvants. However for the four standardised whole-cell actual vaccine formulations (Urovac®, OM-89 or Uro-Vaxom®, Urvakol®, Urostim®), their efficacy was far from convincing [188, 195, 196]. Regarding intestinal pathogenic *E. coli, o*nly one vaccine (Dukoral® produced by SBL Sweden) is currently available for the prevention of ETEC diarrhea. This vaccine has been recommended to prevent travellers' diarrhea in people visiting endemic regions from developed countries [197]. Dukoral® is primarily designed and licensed to prevent diarrhea due to *Vibrio cholera*, but it contains a recombinant B subunit of the cholera toxin that is antigenically very similar to the heat-labile toxin of ETEC [198]. In an early clinical trial, significant cross protection against ETEC diarrhea was demonstrated, using a prototype of this vaccine which contained purified cholera B subunit rather than the recombinant form [199]. Currently, many companies are trying to sort out new protein candidates for vaccines against ETEC, generally those candidates can be categorised in to two groups: i) inactivated vaccines containing killed whole cells, purified colonisation factors antigens, or inactivated heat-labile toxin; ii) live attenuated vaccines containing genetically modified, non-pathogenic strains of ETEC, or alternative carrier bacteria expressing the important ETEC antigens [200].

Despite Dukoral®, no vaccines are currently available to control EPEC infections, especially in hospital environment. Testing purified recombinant versions of EspB and BfpA eliciting an antibody response in rabbits, antigenic potential in humans was shown when reacted with secretory

IgA (sIgA) present in the stools of diarrheic pediatric patients [201, 202]. Also, mice vaccinated with EPEC ghosts show significant protection against lethal challenges [158]. Vaccinated mice showed 84 to 90% protection when challenged with wild-type EPEC, compared to no protection in control mice. Homologous rechallenge with wild-type EPEC resulted in a reduced severity of disease but had no effect on incidence of diarrhea [203].

To date, several studies have been done in order to find a strategy against EHEC with variable success. These works were based on the use of recombinant virulence proteins, e.g. either Stx, intimin and *E. coli* secreted protein A (EspA) [204], peptides [205], fusion proteins Stx2 and Stx1 subunits such as Stx2Am-Stx1B [206], avirulent ghost cells of EHEC O157:H7 [207] or the application of other bacterial species such as *Salmonella* as a carrier for vaccine proteins against mucosal pathogens including EHEC [208]. Recently, DNA vaccination strategies have also been developed. An anti-EHEC DNA-based vaccine utilising the *efa*-1′ gene (EHEC factor for adherence-1) has shown mucosal and systemic immune responses and was able to confer efficient protection against challenge with the EHEC EDL933 in mouse model, suggesting that mucosal inoculation with DNA vaccines as a valid vaccination approach for the induction of immuno-mediated protection against EHEC infections [209]. Also DNA constructs encoding a nontoxic antigenic Stx2 toxoid which induces both specific humoral responses and activity neutralisation of Stx2 [210]. The genomic plasticity of InPEC resulted in versatile species able to colonise, multiply and damaging different environments. The ability *E. coli* virulence factors to affect such a wide range of host cellular functions challenge the effectiveness of a possible vaccine.

1.5. Reverse vaccinology

Since the first reported vaccine more than two centuries ago, vaccine development followed the same basic principles. The official origin of vaccination dates to 1796, when James Phipps was inoculated by Edward Jenner with a bovine poxvirus to induce an immune response protective against the closely related human pathogen smallpox virus. Nearly a century later, Pasteur developed a live attenuated vaccine against rabies and established the following basic steps for vaccine development: isolation, inactivation, and injection of the causative organism. The basic principles postulated by Pasteur and Jenner guided the world vaccine development during the twentieth century, irradicating lethal infectious diseases such diphtheria, tetanus, poliomyelitis, pertussis, measles, mumps, rubella and invasive *Haemophilus influenzae* B [211, 212]. Indeed, all developed vaccines were based in at least one of the following approaches: inactivated

microorganisms, live attenuated and subunit vaccines, including protein-conjugated capsular polysaccharides, toxoids, cell-free extracts, recombinant proteins and stand-alone capsular polysaccharides [212].

With the genome of the first bacterium sequenced in 1995, vaccine development entered in a new era [213]. Suddenly and for the first time in human history, it was made possible to move forward the classical Pasteur's principle and identify vaccine components without the need to grow the disease-causing microorganism but following computer analysis of the coding DNA sequences (CDS) in a bacterial pathogen genome [214]. This new approach named reverse vaccinology aims at developping vaccines starting from the genomic information instead of growing the pathogen [215]. Nowadays, most new vaccine projects apply the concept of reverse vaccinology. This approach is especially relevant when the etiologic agent cannot be grown in the laboratory conditions and/or when the disease is caused by microorganisms that undergo extensive antigenic variation [214]. In theory, it is possible to design a totally synthetic vaccine containing strings of the best epitopes encoded by a microorganism and/or exclude from vaccines those antigens that are predicted to be poorly immunogenic or are potentially cross-reactive with human proteins [215, 216]. As such, reverse vaccinology could allow extending the vaccines development field to new frontiers.

The first pathogen addressed by reverse vaccionology was Meningococcus B (MenB). While the *Neisseria meningitidis* genome sequence was still being assembled, *in silico* analysis allowed the prediction of proteins that could be surface-exposed or homologous to known factors associated with virulence and pathogenesis, leading to the selection of 570 potential vaccine candidates [217]. Those candidates were expressed in *E. coli* and sera from immunized mice were obtained against each of them, resulting in 29 antigens capable to induce antibodies able to kill the bacteria. This revealed the power of the technology once up to date only 5 antigens were described having bactericidal activity [217, 218]. Subsequent long path of vaccine development led to the production of Bexero®, currently on the market.

Since, reverse vaccinology has been then applied to other bacterial pathogens such group B streptococcus [219] and *Chlamydia pneumoniae* [220] but not vaccines are available on the market as yet.

1.6. Proteomics in vaccine development and antigen discovery

Analysis of the genome provides information about genetic background of the cell and potential regulatory mechanisms. Contrasting with the genome, transcriptome and the proteome are dynamic and reflect the adaptative capabilities of a living cell to different environmental conditions and stimuli. However, transcriptomic data are rarely consistent with that generated from proteomic analysis, implying a role of post-transcriptional mechanisms that modulate the cell proteome profile [221]. Indeed, the expression of a gene at transcript level does not necessarily correlate with protein translation and can even be inversely proportionate. In fact, the all transcripts correspond is a pool of mRNA, which actually results from two processes, their biosynthesis following DNA transcription but also their degradation, which is often a phenomenon quite overlooked. Besides, each mRNA has a different half-life, which will consequently affect RNA translation and ultimately protein expression. Undeniably, genetic, epigenetic, transcriptional and post-transcriptional regulations play a key role in protein expression [222]. When it comes to global protein expression, proteomic approaches are the methods of choice [222, 223]. Over the last years bioinformatics analysis of proteome data has led to the definition of several new "omes" relevant to antigen discovery, such as the exoproteome (extracellular proteome), proteosurfaceome (cell-surface proteome), or interactome (protein-protein interaction) [224-226]. The rapid development of proteomic strategies such protein crystallisation methods, magnetic mesonance imaging (MRI), mass spectrometry technologies, cell fractionation procedures coupled with specialized and powered analytical software have facilitated antigen discovery. However, the analysis of proteins from cell envelope has proven difficult. The different physicochemical properties, dynamic range of abundances, subcellular localisations, and binding partners of proteins established an big challenge to proteomic research [227]. The cell-surface proteome of several pathogens is now accessed, by three different methods, namely (i) cell fractionation [228], (ii) proteolytic shaving [229], (iii) biotinylation [230].

Subproteome fractionation is the classical and oldest method for membrane proteins enrichment. This method involves subjecting cell lysates to differential centrifugation, wherein cellular components are separated on the basis of particle size and shape according to the density and viscosity of the gradient medium [231]. However, recurrent limitations of these fractionation-based enrichment strategy include the need for large amounts of starting biomass material and the level of cross-contamination obtained due to the presence of abundant proteins derived from other subcellular locations [232].

Cell shaving procedure involves the proteolytic digestion of surface exposed proteins of intact cells with proteases that are commonly used in MS analysis such trypsin or protease K. Generally, that digestion must be performed for short periods of time and cells must be kept under isotonic conditions that limit cell rupture or lysis, to ensure recovering surface exposed proteins and epitopes [227]. While this trypsin shaving approach can theatrically be applied to Gram-negative and Gram-positive bacteria alike to examine their cell surface proteome, in practice this proves difficult. This is mainly due to the relative fragility of Gram-negative cell envelope, which leads to increased cell lysis and thus identification of more false-positive cytoplasmic proteins [233].

The lastest methods involve the use of molecular tags to label cell surface-exposed proteins. Labeling technique is performed on intact cells, mainly using fluorescent or biotin reagents to bind surface-exposed proteins. Ideally, these tagging reagents should not be able to enter and bind inner components of the cell. Protein fractions are then recovered and the tagged proteins purified by affinity chromatography [226, 230]. A major advantage of cell surface labelling over cell shaving method is that it can be applied to the Gram-positive and Gram-negative bacterial cell envelope alike since it does not compromise the cellular integrity [227, 234]. Regarding biotinylation reagents, their low molecular weight does not appear to modify the proteins conformational structure and it has high specificity for avidin, enabling an easy and selective purification method [235].

For proteomic analysis per se, two main approaches can be applied, namely those related with bidimensional polyacrylamide gel electrophoresis (2-D PAGE) and off-gel approaches [236]. 2-D PAGE set the gold standard for protein separation for years. By greatly improving the resolution of complex protein samples compared to monodimensional polyacrylamide gel electrophoresis (1-D PAGE), this approach allows identifying a few hundred different proteins when combined with mass spectrometry (MS) analysis based on mass profile fingerprints performed from separated spots. The sizes of the spots offer semi-quantitative analysis and the visualisation of protein isoforms, which required to be all identified for investigating the differential expression of a given protein between two experimental conditions [232, 236]. Another intrinsic limit to this approach is the dynamic range whenever the isofocalisation (pI) in the first dimension, the size of the protein in the second dimension and also the soluble nature of the proteins; in other words, proteins with extreme pI and size and/or hydrophobicity are missed when performing a 2-D PAGE. In other hand, off-gel proteomics (also known as shotgun proteomics or multidimentional protein identification technology) has been proven a powerful high-throughput tool to investigate different proteomes. Because proteins in a sample are directly subjected to trypsic digestion, there is no more issue

relative pI and size and/or hydrophobicity. The nature and quantity of protein recover will mainly depend on the initial protein extraction protocol, whereas the mixture of trypsic peptides is separated and subsequently analysed by MS. Pre-fractionation of the proteins prior to the proteolysis step or analysis of sub-proteomes can be advantageous [227, 236]. Proteomics is a complementary approaches to genomics and transcriptomics but also much more appropriate and useful in identifying surface proteins. These high-throughput molecular profiling approaches have also indirectly contributed to the field of vaccinology and/or improving our understanding of pathogenesis and host–pathogen interactions.

REFERENCES

1. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nature reviews Microbiology* 2010; **8**: 207-17.
2. Hooper LV, Gordon JI. Commensal host-bacterial relationships in the gut. *Science* 2001; **292**: 1115-8.
3. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. *Nature reviews Microbiology* 2004; **2**: 123-40.
4. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clinical microbiology reviews* 1998; **11**: 142-201.
5. Karch H, Tarr PI, Bielaszewska M. Enterohaemorrhagic *Escherichia coli* in human medicine. *International journal of medical microbiology : IJMM* 2005; **295**: 405-18.
6. Cohen MB, Nataro JP, Bernstein DI et al. Prevalence of diarrheagenic *Escherichia coli* in acute childhood enteritis: a prospective controlled study. *The Journal of pediatrics* 2005; **146**: 54-61.
7. Schmidt MA. LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* 2010; **12**: 1544-52.
8. Dobrindt U. (Patho-)Genomics of Escherichia coli. *International journal of medical microbiology : IJMM* 2005; **295**: 357-71.
9. Nataro JP, Mai V, Johnson J et al. Diarrheagenic *Escherichia coli* infection in Baltimore, Maryland, and New Haven, Connecticut. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006; **43**: 402-7.
10. Creuzburg K, Heeren S, Lis CM et al. Genetic background and mobility of variants of the gene *nleA* in attaching and effacing *Escherichia coli*. *Applied and environmental microbiology* 2011; **77**: 8705-13.
11. Creuzburg K, Middendorf B, Mellmann A et al. Evolutionary analysis and distribution of type III effector genes in pathogenic *Escherichia coli* from human, animal and food sources. *Environmental microbiology* 2011; **13**: 439-52.
12. Qadri F, Svennerholm AM, Faruque AS, Sack RB. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clinical microbiology reviews* 2005; **18**: 465-83.
13. Amisano G, Fornasero S, Migliaretti G et al. Diarrheagenic *Escherichia coli* in acute gastroenteritis in infants in North-West Italy. *The new microbiologica* 2011; **34**: 45-51.
14. McDonough MA, Butterton JR. Spontaneous tandem amplification and deletion of the shiga toxin operon in *Shigella dysenteriae* 1. *Molecular microbiology* 1999; **34**: 1058-69.
15. Chaudhuri RR, Henderson IR. The evolution of the *Escherichia coli* phylogeny. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2012; **12**: 214-26.
16. Lan R, Alles MC, Donohoe K et al. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. *Infection and immunity* 2004; **72**: 5080-8.
17. Scheutz F, Teel LD, Beutin L et al. Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *Journal of clinical microbiology* 2012; **50**: 2951-63.
18. Calderwood S, Acheson DWK, Keusch GT et al. Proposed new nomenclature for SLT (VT) family. *ASM News* 1996; **62**: 118-9.
19. Melton-Celsa AR, O'Brien AD. Shiga Toxins of *Shigella dysenteriae* and *Escherichia coli*. In: Aktories K, Just I, eds. *Bacterial Protein Toxins*: Springer Berlin Heidelberg 2000:385-406.

20. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 2005; **365**: 1073-86.
21. Walker CL, Black RE. Diarrhoea morbidity and mortality in older children, adolescents, and adults. *Epidemiology and infection* 2010; **138**: 1215-26.
22. Karch H, Bielaszewska M, Bitzan M, Schmidt H. Epidemiology and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Diagnostic microbiology and infectious disease* 1999; **34**: 229-43.
23. Kelly JK, Pai CH, Jadusingh IH et al. The histopathology of rectosigmoid biopsies from adults with bloody diarrhea due to verotoxin-producing *Escherichia coli*. *American journal of clinical pathology* 1987; **88**: 78-82.
24. Lewis SB, Cook V, Tighe R, Schuller S. Enterohemorrhagic *Escherichia coli* colonization of human colonic epithelium *in vitro* and *ex vivo*. *Infection and immunity* 2015; **83**: 942-9.
25. Karmali MA, Gannon V, Sargeant JM. Verocytotoxin-producing *Escherichia coli* (VTEC). *Veterinary microbiology* 2010; **140**: 360-70.
26. Karmali MA, Mascarenhas M, Shen S et al. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. *Journal of clinical microbiology* 2003; **41**: 4930-40.
27. Mora A, Blanco M, Blanco JE et al. Serotypes, virulence genes and intimin types of Shiga toxin (verocytotoxin)-producing *Escherichia coli* isolates from minced beef in Lugo (Spain) from 1995 through 2003. *BMC microbiology* 2007; **7**: 13.
28. Frohlicher E, Krause G, Zweifel C et al. Characterization of attaching and effacing *Escherichia coli* (AEEC) isolated from pigs and sheep. *BMC microbiology* 2008; **8**: 144.
29. Kyle JL, Cummings CA, Parker CT et al. *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. *Journal of bacteriology* 2012; **194**: 1885-96.
30. Caprioli A, Morabito S, Brugere H, Oswald E. Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Veterinary research* 2005; **36**: 289-311.
31. Hauser E, Mellmann A, Semmler T et al. Phylogenetic and molecular analysis of food-borne shiga toxin-producing *Escherichia coli*. *Applied and environmental microbiology* 2013; **79**: 2731-40.
32. Brzuszkiewicz E, Thurmer A, Schuldes J et al. Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic *Escherichia coli* (EAHEC). *Archives of microbiology* 2011; **193**: 883-91.
33. Mellmann A, Harmsen D, Cummings CA et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PloS one* 2011; **6**: e22751.
34. Jones TF, Sashti N, Ingram A et al. Characteristics of Clusters of *Salmonella* and *Escherichia coli* O157 Detected by Pulsed-Field Gel Electrophoresis that Predict Identification of Outbreaks. *Foodborne Pathog Dis* 2016; **13**: 674-8.
35. Ho NK, Henry AC, Johnson-Henry K, Sherman PM. Pathogenicity, host responses and implications for management of enterohemorrhagic *Escherichia coli* O157:H7 infection. *Can J Gastroenterol* 2013; **27**: 281-5.
36. Bavaro MF. *E. coli* O157:H7 and other toxigenic strains: the curse of global food distribution. *Curr Gastroenterol Rep* 2012; **14**: 317-23.
37. Chromek M, Arvidsson I, Karpman D. The antimicrobial peptide cathelicidin protects mice from *Escherichia coli* O157:H7-mediated disease. *PloS one* 2012; **7**: e46476.

38. Barnett Foster D. Modulation of the enterohemorrhagic *E. coli* virulence program through the human gastrointestinal tract. *Virulence* 2013; **4**: 315-23.
39. McKee ML, O'Brien AD. Truncated enterohemorrhagic *Escherichia coli* (EHEC) O157:H7 intimin (EaeA) fusion proteins promote adherence of EHEC strains to HEp-2 cells. *Infection and immunity* 1996; **64**: 2225-33.
40. Herold S, Paton JC, Srimanote P, Paton AW. Differential effects of short-chain fatty acids and iron on expression of iha in Shiga-toxigenic *Escherichia coli*. *Microbiology* 2009; **155**: 3554-63.
41. Hughes DT, Clarke MB, Yamamoto K et al. The QseC adrenergic signaling cascade in Enterohemorrhagic *E. coli* (EHEC). *PLoS Pathog* 2009; **5**: e1000553.
42. DeVinney R, Gauthier A, Abe A, Finlay BB. Enteropathogenic *Escherichia coli*: a pathogen that inserts its own receptor into host cells. *Cell Mol Life Sci* 1999; **55**: 961-76.
43. McNamara BP, Donnenberg MS. A novel proline-rich protein, EspF, is secreted from enteropathogenic *Escherichia coli* via the type III export pathway. *FEMS Microbiol Lett* 1998; **166**: 71-8.
44. Kenny B, DeVinney R, Stein M et al. Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* 1997; **91**: 511-20.
45. Campellone KG, Robbins D, Leong JM. EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. *Dev Cell* 2004; **7**: 217-28.
46. Weiss SM, Ladwein M, Schmidt D et al. IRSp53 links the enterohemorrhagic *E. coli* effectors Tir and EspFU for actin pedestal formation. *Cell Host Microbe* 2009; **5**: 244-58.
47. Trinchieri G, Sher A. Cooperation of Toll-like receptor signals in innate immune defence. *Nat Rev Immunol* 2007; **7**: 179-90.
48. Kastenmuller W, Torabi-Parizi P, Subramanian N et al. A spatially-organized multicellular innate immune response in lymph nodes limits systemic pathogen spread. *Cell* 2012; **150**: 1235-48.
49. Shea-Donohue T, Fasano A, Smith A, Zhao A. Enteric pathogens and gut function: Role of cytokines and STATs. *Gut Microbes* 2010; **1**: 316-24.
50. Casanova JL, Holland SM, Notarangelo LD. Inborn errors of human JAKs and STATs. *Immunity* 2012; **36**: 515-28.
51. Stoiber D, Stockinger S, Steinlein P et al. *Listeria monocytogenes* modulates macrophage cytokine responses through STAT serine phosphorylation and the induction of suppressor of cytokine signaling 3. *J Immunol* 2001; **166**: 466-72.
52. Ho NK, Ossa JC, Silphaduang U et al. Enterohemorrhagic *Escherichia coli* O157:H7 Shiga toxins inhibit gamma interferon-mediated cellular activation. *Infection and immunity* 2012; **80**: 2307-15.
53. Finlay BB, Falkow S. Common themes in microbial pathogenicity revisited. *Microbiology and molecular biology reviews : MMBR* 1997; **61**: 136-69.
54. Chagnot C, Zorgani MA, Astruc T, Desvaux M. Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front Microbiol* 2013; **4**: 303.
55. Tjalsma H, Bolhuis A, Jongbloed JD et al. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiology and molecular biology reviews : MMBR* 2000; **64**: 515-47.
56. Antelmann H, Tjalsma H, Voigt B et al. A proteomic view on genome-based signal peptide predictions. *Genome research* 2001; **11**: 1484-502.
57. Paton AW, Srimanote P, Woodrow MC, Paton JC. Characterization of Saa, a novel autoagglutinating adhesin produced by *locus* of enterocyte effacement-negative Shiga-

toxigenic *Escherichia coli* strains that are virulent for humans. *Infection and immunity* 2001; **69**: 6999-7009.
58. Gerlach RG, Hensel M. Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *International journal of medical microbiology : IJMM* 2007; **297**: 401-15.
59. Knowles TJ, Scott-Tucker A, Overduin M, Henderson IR. Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nature reviews Microbiology* 2009; **7**: 206-14.
60. Tarr PI, Bilge SS, Vary JC, Jr. et al. Iha: a novel *Escherichia coli* O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure. *Infection and immunity* 2000; **68**: 1400-7.
61. Torres AG, Kaper JB. Multiple elements controlling adherence of enterohemorrhagic *Escherichia coli* O157:H7 to HeLa cells. *Infection and immunity* 2003; **71**: 4985-95.
62. Monteiro R, Ageorges V, Rojas-Lopez M et al. A secretome view of colonisation factors in Shiga toxin-encoding *Escherichia coli* (STEC): from enterohaemorrhagic E. coli (EHEC) to related enteropathotypes. *FEMS Microbiol Lett* 2016; **363**.
63. Thomas S, Holland IB, Schmitt L. The Type 1 secretion pathway - the hemolysin system and beyond. *Biochim Biophys Acta* 2014; **1843**: 1629-41.
64. Nishi J, Sheikh J, Mizuguchi K et al. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. *J Biol Chem* 2003; **278**: 45680-9.
65. Velarde JJ, Varney KM, Inman KG et al. Solution structure of the novel dispersin protein of enteroaggregative *Escherichia coli*. *Molecular microbiology* 2007; **66**: 1123-35.
66. Sheikh J, Czeczulin JR, Harrington S et al. A novel dispersin protein in enteroaggregative *Escherichia coli*. *J Clin Invest* 2002; **110**: 1329-37.
67. Monteiro BT, Campos LC, Sircili MP et al. The dispersin-encoding gene (aap) is not restricted to enteroaggregative *Escherichia coli*. *Diagnostic microbiology and infectious disease* 2009; **65**: 81-4.
68. Muniesa M, Hammerl JA, Hertwig S et al. Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Applied and environmental microbiology* 2012; **78**: 4065-73.
69. Voulhoux R, Ball G, Ize B et al. Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *The EMBO journal* 2001; **20**: 6735-41.
70. Durand E, Alphonse S, Brochier-Armanet C et al. The assembly mode of the pseudopilus: a hallmark to distinguish a novel secretion system subtype. *J Biol Chem* 2011; **286**: 24407-16.
71. Ayers M, Howell PL, Burrows LL. Architecture of the type II secretion and type IV pilus machineries. *Future Microbiol* 2010; **5**: 1203-18.
72. Craig L, Pique ME, Tainer JA. Type IV pilus structure and bacterial pathogenicity. *Nature reviews Microbiology* 2004; **2**: 363-78.
73. Giron JA, Ho AS, Schoolnik GK. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. *Science* 1991; **254**: 710-3.
74. Scaletsky IC, Silva ML, Trabulsi LR. Distinctive patterns of adherence of enteropathogenic *Escherichia coli* to HeLa cells. *Infection and immunity* 1984; **45**: 534-6.
75. Stone KD, Zhang HZ, Carlson LK, Donnenberg MS. A cluster of fourteen genes from enteropathogenic *Escherichia coli* is sufficient for the biogenesis of a type IV pilus. *Molecular microbiology* 1996; **20**: 325-37.
76. Giron JA, Donnenberg MS, Martin WC et al. Distribution of the bundle-forming pilus structural gene (bfpA) among enteropathogenic *Escherichia coli*. *J Infect Dis* 1993; **168**: 1037-41.

77. Croxen MA, Law RJ, Scholz R et al. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clinical microbiology reviews* 2013; **26**: 822-80.
78. Xicohtencatl-Cortes J, Monteiro-Neto V, Saldana Z et al. The type 4 pili of enterohemorrhagic *Escherichia coli* O157:H7 are multipurpose structures with pathogenic attributes. *Journal of bacteriology* 2009; **191**: 411-21.
79. Dudley EG, Abe C, Ghigo JM et al. An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative Escherichia coli strain C1096 to cultured cells and abiotic surfaces. *Infection and immunity* 2006; **74**: 2102-14.
80. Diepold A, Armitage JP. Type III secretion systems: the bacterial flagellum and the injectisome. *Philos Trans R Soc Lond B Biol Sci* 2015; **370**.
81. Desvaux M, Hebraud M, Henderson IR, Pallen MJ. Type III secretion: what's in a name? *Trends Microbiol* 2006; **14**: 157-60.
82. Galan JE, Wolf-Watz H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* 2006; **444**: 567-73.
83. Cornelis GR. The type III secretion injectisome. *Nature reviews Microbiology* 2006; **4**: 811-25.
84. Wong AR, Pearson JS, Bright MD et al. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Molecular microbiology* 2011; **80**: 1420-38.
85. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* 1995; **92**: 1664-8.
86. Donnenberg MS, Tacket CO, James SP et al. Role of the eaeA gene in experimental enteropathogenic *Escherichia coli* infection. *J Clin Invest* 1993; **92**: 1412-7.
87. Liu H, Magoun L, Luperchio S et al. The Tir-binding region of enterohaemorrhagic *Escherichia coli* intimin is sufficient to trigger actin condensation after bacterial-induced host cell signalling. *Molecular microbiology* 1999; **34**: 67-81.
88. Shaw RK, Berger CN, Feys B et al. Enterohemorrhagic *Escherichia coli* exploits EspA filaments for attachment to salad leaves. *Applied and environmental microbiology* 2008; **74**: 2908-14.
89. Berger CN, Sodha SV, Shaw RK et al. Fresh fruit and vegetables as vehicles for the transmission of human pathogens. *Environmental microbiology* 2010; **12**: 2385-97.
90. Schroeder GN, Hilbi H. Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clinical microbiology reviews* 2008; **21**: 134-56.
91. DiLuzio WR, Turner L, Mayer M et al. *Escherichia coli* swim on the right-hand side. *Nature* 2005; **435**: 1271-4.
92. Altegoer F, Bange G. Undiscovered regions on the molecular landscape of flagellar assembly. *Curr Opin Microbiol* 2015; **28**: 98-105.
93. Macnab RM. Type III flagellar protein export and flagellar assembly. *Biochim Biophys Acta* 2004; **1694**: 207-17.
94. Turner L, Zhang R, Darnton NC, Berg HC. Visualization of Flagella during bacterial Swarming. *Journal of bacteriology* 2010; **192**: 3259-67.
95. Haiko J, Westerlund-Wikstrom B. The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel)* 2013; **2**: 1242-67.
96. Mahajan A, Currie CG, Mackie S et al. An investigation of the expression and adhesin function of H7 flagella in the interaction of *Escherichia coli* O157 : H7 with bovine intestinal epithelium. *Cell Microbiol* 2009; **11**: 121-37.

97. McNeilly TN, Naylor SW, Mahajan A et al. *Escherichia coli* O157:H7 colonization in cattle following systemic and mucosal immunization with purified H7 flagellin. *Infection and immunity* 2008; **76**: 2594-602.
98. Alvarez-Martinez CE, Christie PJ. Biological diversity of prokaryotic type IV secretion systems. *Microbiology and molecular biology reviews : MMBR* 2009; **73**: 775-808.
99. Christie PJ, Whitaker N, Gonzalez-Rivera C. Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta* 2014; **1843**: 1578-91.
100. Christie PJ, Vogel JP. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* 2000; **8**: 354-60.
101. Lawley TD, Klimke WA, Gubbins MJ, Frost LS. F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* 2003; **224**: 1-15.
102. Lim JY, Yoon J, Hovde CJ. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J Microbiol Biotechnol* 2010; **20**: 5-14.
103. Ostroff SM, Tarr PI, Neill MA et al. Toxin genotypes and plasmid profiles as determinants of systemic sequelae in *Escherichia coli* O157:H7 infections. *J Infect Dis* 1989; **160**: 994-8.
104. Paros M, Tarr PI, Kim H et al. A comparison of human and bovine *Escherichia coli* O157:H7 isolates by toxin genotype, plasmid profile, and bacteriophage lambda-restriction fragment length polymorphism profile. *J Infect Dis* 1993; **168**: 1300-3.
105. Wang P, Xiong Y, Lan R et al. pO157_Sal, a novel conjugative plasmid detected in outbreak isolates of *Escherichia coli* O157:H7. *Journal of clinical microbiology* 2011; **49**: 1594-7.
106. Fratamico PM, Yan X, Caprioli A et al. The complete DNA sequence and analysis of the virulence plasmid and of five additional plasmids carried by Shiga toxin-producing *Escherichia coli* O26:H11 strain H30. *International journal of medical microbiology : IJMM* 2011; **301**: 192-203.
107. Ghigo JM. Natural conjugative plasmids induce bacterial biofilm development. *Nature* 2001; **412**: 442-5.
108. May T, Okabe S. *Escherichia coli* harboring a natural IncF conjugative F plasmid develops complex mature biofilms by stimulating synthesis of colanic acid and Curli. *Journal of bacteriology* 2008; **190**: 7479-90.
109. May T, Ito A, Okabe S. Characterization and global gene expression of F- phenocopies during *Escherichia coli* biofilm formation. *Mol Genet Genomics* 2010; **284**: 333-42.
110. Choi PS, Dawson AJ, Bernstein HD. Characterization of a novel two-partner secretion system in *Escherichia coli* O157:H7. *Journal of bacteriology* 2007; **189**: 3452-61.
111. Desvaux M, Parham NJ, Henderson IR. Type V protein secretion: simplicity gone awry? *Current issues in molecular biology* 2004; **6**: 111-24.
112. Benz I, Schmidt MA. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infection and immunity* 1989; **57**: 1506-11.
113. Klemm P, Vejborg RM, Sherlock O. Self-associating autotransporters, SAATs: functional and structural similarities. *International journal of medical microbiology : IJMM* 2006; **296**: 187-95.
114. Sherlock O, Schembri MA, Reisner A, Klemm P. Novel roles for the AIDA adhesin from diarrheagenic *Escherichia coli*: cell aggregation and biofilm formation. *Journal of bacteriology* 2004; **186**: 8058-65.
115. Wells TJ, Sherlock O, Rivas L et al. EhaA is a novel autotransporter protein of enterohemorrhagic *Escherichia coli* O157:H7 that contributes to adhesion and biofilm formation. *Environmental microbiology* 2008; **10**: 589-604.

116. Wells TJ, McNeilly TN, Totsika M et al. The *Escherichia coli* O157:H7 EhaB autotransporter protein binds to laminin and collagen I and induces a serum IgA response in O157:H7 challenged cattle. *Environmental microbiology* 2009; **11**: 1803-14.
117. Easton DM, Totsika M, Allsopp LP et al. Characterization of EhaJ, a New Autotransporter Protein from Enterohemorrhagic and Enteropathogenic *Escherichia coli*. *Front Microbiol* 2011; **2**: 120.
118. Torres AG, Jeter C, Langley W, Matthysse AG. Differential binding of *Escherichia coli* O157:H7 to alfalfa, human epithelial cells, and plastic is mediated by a variety of surface structures. *Applied and environmental microbiology* 2005; **71**: 8008-15.
119. Henderson IR, Navarro-Garcia F, Desvaux M et al. Type V protein secretion pathway: the autotransporter story. *Microbiology and molecular biology reviews : MMBR* 2004; **68**: 692-744.
120. Dziva F, Mahajan A, Cameron P et al. EspP, a Type V-secreted serine protease of enterohaemorrhagic *Escherichia coli* O157:H7, influences intestinal colonization of calves and adherence to bovine primary intestinal epithelial cells. *FEMS Microbiol Lett* 2007; **271**: 258-64.
121. Puttamreddy S, Cornick NA, Minion FC. Genome-wide transposon mutagenesis reveals a role for pO157 genes in biofilm development in *Escherichia coli* O157:H7 EDL933. *Infection and immunity* 2010; **78**: 2377-84.
122. Xicohtencatl-Cortes J, Saldana Z, Deng W et al. Bacterial macroscopic rope-like fibers with cytopathic and adhesive properties. *J Biol Chem* 2010; **285**: 32336-42.
123. van Diemen PM, Dziva F, Stevens MP, Wallis TS. Identification of enterohemorrhagic *Escherichia coli* O26:H- genes required for intestinal colonization in calves. *Infection and immunity* 2005; **73**: 1735-43.
124. Herold S, Paton JC, Paton AW. Sab, a novel autotransporter of locus of enterocyte effacement-negative shiga-toxigenic *Escherichia coli* O113:H21, contributes to adherence and biofilm formation. *Infection and immunity* 2009; **77**: 3234-43.
125. Szczesny P, Lupas A. Domain annotation of trimeric autotransporter adhesins--daTAA. *Bioinformatics* 2008; **24**: 1251-6.
126. Linke D, Riess T, Autenrieth IB et al. Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol* 2006; **14**: 264-70.
127. El Tahir Y, Skurnik M. YadA, the multifaceted *Yersinia* adhesin. *International journal of medical microbiology : IJMM* 2001; **291**: 209-18.
128. Totsika M, Wells TJ, Beloin C et al. Molecular characterization of the EhaG and UpaG trimeric autotransporter proteins from pathogenic *Escherichia coli*. *Applied and environmental microbiology* 2012; **78**: 2179-89.
129. Lu Y, Iyoda S, Satou H et al. A new immunoglobulin-binding protein, EibG, is responsible for the chain-like adhesion phenotype of locus of enterocyte effacement-negative, shiga toxin-producing *Escherichia coli*. *Infection and immunity* 2006; **74**: 5747-55.
130. Merkel V, Ohder B, Bielaszewska M et al. Distribution and phylogeny of immunoglobulin-binding protein G in Shiga toxin-producing *Escherichia coli* and its association with adherence phenotypes. *Infection and immunity* 2010; **78**: 3625-36.
131. Leo JC, Oberhettinger P, Schutz M, Linke D. The inverse autotransporter family: intimin, invasin and related proteins. *International journal of medical microbiology : IJMM* 2015; **305**: 276-82.
132. Oberhettinger P, Schutz M, Leo JC et al. Intimin and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared to classical autotransport. *PloS one* 2012; **7**: e47069.

133. Oberhettinger P, Leo JC, Linke D et al. The inverse autotransporter intimin exports its passenger domain via a hairpin intermediate. *J Biol Chem* 2015; **290**: 1837-49.
134. Adu-Bobie J, Frankel G, Bain C et al. Detection of intimins alpha, beta, gamma, and delta, four intimin derivatives expressed by attaching and effacing microbial pathogens. *Journal of clinical microbiology* 1998; **36**: 662-8.
135. Oswald E, Schmidt H, Morabito S et al. Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli:* characterization of a new intimin variant. *Infection and immunity* 2000; **68**: 64-71.
136. Mallick EM, Brady MJ, Luperchio SA et al. Allele- and tir-independent functions of intimin in diverse animal infection models. *Front Microbiol* 2012; **3**: 11.
137. Liu H, Magoun L, Leong JM. beta1-chain integrins are not essential for intimin-mediated host cell attachment and enteropathogenic *Escherichia coli*-induced actin condensation. *Infection and immunity* 1999; **67**: 2045-9.
138. Nesta B, Spraggon G, Alteri C et al. FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *MBio* 2012; **3**.
139. Easton DM, Allsopp LP, Phan MD et al. The Intimin-Like Protein FdeC Is Regulated by H-NS and Temperature in Enterohemorrhagic *Escherichia coli*. *Applied and environmental microbiology* 2014; **80**: 7337-47.
140. Leiman PG, Basler M, Ramagopal UA et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* 2009; **106**: 4154-9.
141. Aschtgen MS, Gavioli M, Dessen A et al. The SciZ protein anchors the enteroaggregative *Escherichia coli* Type VI secretion system to the cell wall. *Molecular microbiology* 2010; **75**: 886-99.
142. Lloyd AL, Henderson TA, Vigil PD, Mobley HL. Genomic islands of uropathogenic *Escherichia coli* contribute to virulence. *Journal of bacteriology* 2009; **191**: 3469-81.
143. Moriel DG, Bertoldi I, Spagnuolo A et al. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 2010; **107**: 9072-7.
144. Shrivastava S, Mande SS. Identification and functional characterization of gene components of Type VI Secretion system in bacterial genomes. *PloS one* 2008; **3**: e2955.
145. Dudley EG, Thomson NR, Parkhill J et al. Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative *Escherichia coli*. *Molecular microbiology* 2006; **61**: 1267-82.
146. Aschtgen MS, Bernard CS, De Bentzmann S et al. SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*. *Journal of bacteriology* 2008; **190**: 7523-31.
147. Desvaux M, Hebraud M, Talon R, Henderson IR. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol* 2009; **17**: 139-45.
148. Busch A, Waksman G. Chaperone-usher pathways: diversity and pilus assembly mechanism. *Philos Trans R Soc Lond B Biol Sci* 2012; **367**: 1112-22.
149. Thanassi DG, Bliska JB, Christie PJ. Surface organelles assembled by secretion systems of Gram-negative bacteria: diversity in structure and function. *FEMS microbiology reviews* 2012; **36**: 1046-82.
150. Zav'yalov V, Zavialov A, Zav'yalova G, Korpela T. Adhesive organelles of Gram-negative pathogens assembled with the classical chaperone/usher machinery: structure and function from a clinical standpoint. *FEMS microbiology reviews* 2010; **34**: 317-78.

151. Hayashi T, Makino K, Ohnishi M et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 2001; **8**: 11-22.
152. Korea CG, Badouraly R, Prevost MC et al. *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environmental microbiology* 2010; **12**: 1957-77.
153. Low AS, Holden N, Rosser T et al. Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. *Environmental microbiology* 2006; **8**: 1033-47.
154. Saldana Z, Erdem AL, Schuller S et al. The *Escherichia coli* common pilus and the bundle-forming pilus act in concert during the formation of localized adherence by enteropathogenic *E. coli*. *Journal of bacteriology* 2009; **191**: 3451-61.
155. Rendon MA, Saldana Z, Erdem AL et al. Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization. *Proc Natl Acad Sci U S A* 2007; **104**: 10637-42.
156. Rossez Y, Holmes A, Lodberg-Pedersen H et al. *Escherichia coli* common pilus (ECP) targets arabinosyl residues in plant cell walls to mediate adhesion to fresh produce plants. *J Biol Chem* 2014; **289**: 34349-65.
157. Doughty S, Sloan J, Bennett-Wood V et al. Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of enterocyte effacement-negative strains of enterohemorrhagic *Escherichia coli*. *Infection and immunity* 2002; **70**: 6761-9.
158. Farfan MJ, Torres AG. Molecular mechanisms that mediate colonization of Shiga toxin-producing *Escherichia coli* strains. *Infection and immunity* 2012; **80**: 903-13.
159. Farfan MJ, Cantero L, Vidal R et al. Long polar fimbriae of enterohemorrhagic *Escherichia coli* O157:H7 bind to extracellular matrix proteins. *Infection and immunity* 2011; **79**: 3744-50.
160. Torres AG, Slater TM, Patel SD et al. Contribution of the Ler- and H-NS-regulated long polar fimbriae of *Escherichia coli* O157:H7 during binding to tissue-cultured cells. *Infection and immunity* 2008; **76**: 5062-71.
161. Ross BN, Rojas-Lopez M, Cieza RJ et al. The Role of Long Polar Fimbriae in *Escherichia coli* O104:H4 Adhesion and Colonization. *PloS one* 2015; **10**: e0141845.
162. Samadder P, Xicohtencatl-Cortes J, Saldana Z et al. The *Escherichia coli* ycbQRST operon encodes fimbriae with laminin-binding and epithelial cell adherence properties in Shiga-toxigenic *E. coli* O157:H7. *Environmental microbiology* 2009; **11**: 1815-26.
163. Musken A, Bielaszewska M, Greune L et al. Anaerobic conditions promote expression of Sfp fimbriae and adherence of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM to human intestinal epithelial cells. *Applied and environmental microbiology* 2008; **74**: 1087-93.
164. Low AS, Dziva F, Torres AG et al. Cloning, expression, and characterization of fimbrial operon F9 from enterohemorrhagic *Escherichia coli* O157:H7. *Infection and immunity* 2006; **74**: 2233-44.
165. Li B, Koch WH, Cebula TA. Detection and characterization of the fimA gene of *Escherichia coli* O157:H7. *Molecular and cellular probes* 1997; **11**: 397-406.
166. Enami M, Nakasone N, Honma Y et al. Expression of type I pili is abolished in verotoxin-producing *Escherichia coli* O157. *FEMS Microbiol Lett* 1999; **179**: 467-72.
167. Galfi P, Neogrady S, Semjen G et al. Attachment of different *Escherichia coli* strains to cultured rumen epithelial cells. *Veterinary microbiology* 1998; **61**: 191-7.

168. Cookson AL, Cooley WA, Woodward MJ. The role of type 1 and curli fimbriae of Shiga toxin-producing *Escherichia coli* in adherence to abiotic surfaces. *International journal of medical microbiology : IJMM* 2002; **292**: 195-205.
169. Roe AJ, Currie C, Smith DG, Gally DL. Analysis of type 1 fimbriae expression in verotoxigenic *Escherichia coli*: a comparison between serotypes O157 and O26. *Microbiology* 2001; **147**: 145-52.
170. Tietze E, Dabrowski PW, Prager R et al. Comparative genomic analysis of two novel sporadic Shiga toxin-producing *Escherichia coli* O104:H4 strains isolated 2011 in Germany. *PloS one* 2015; **10**: e0122074.
171. Boisen N, Struve C, Scheutz F et al. New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family. *Infection and immunity* 2008; **76**: 3281-92.
172. Farfan MJ, Inman KG, Nataro JP. The major pilin subunit of the AAF/II fimbriae from enteroaggregative *Escherichia coli* mediates binding to extracellular matrix proteins. *Infection and immunity* 2008; **76**: 4378-84.
173. Boisen N, Melton-Celsa AR, Scheutz F et al. Shiga toxin 2a and Enteroaggregative *Escherichia coli*--a deadly combination. *Gut Microbes* 2015; **6**: 272-8.
174. Gonyar LA, Kendall MM. Ethanolamine and choline promote expression of putative and characterized fimbriae in enterohemorrhagic *Escherichia coli* O157:H7. *Infection and immunity* 2014; **82**: 193-201.
175. Goyal P, Krasteva PV, Van Gerven N et al. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* 2014; **516**: 250-3.
176. Hammar M, Bian Z, Normark S. Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli*. *Proc Natl Acad Sci U S A* 1996; **93**: 6562-6.
177. Fronzes R, Christie PJ, Waksman G. The structural biology of type IV secretion systems. *Nature reviews Microbiology* 2009; **7**: 703-14.
178. Uhlich GA, Keen JE, Elder RO. Variations in the csgD promoter of *Escherichia coli* O157:H7 associated with increased virulence in mice and increased invasion of HEp-2 cells. *Infection and immunity* 2002; **70**: 395-9.
179. Lloyd SJ, Ritchie JM, Rojas-Lopez M et al. A double, long polar fimbria mutant of *Escherichia coli* O157:H7 expresses Curli and exhibits reduced in vivo colonization. *Infection and immunity* 2012; **80**: 914-20.
180. Ryu JH, Beuchat LR. Biofilm formation by *Escherichia coli* O157:H7 on stainless steel: effect of exopolysaccharide and Curli production on its resistance to chlorine. *Applied and environmental microbiology* 2005; **71**: 247-54.
181. Henderson B, Martin A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infection and immunity* 2011; **79**: 3476-91.
182. Egea L, Aguilera L, Gimenez R et al. Role of secreted glyceraldehyde-3-phosphate dehydrogenase in the infection mechanism of enterohemorrhagic and enteropathogenic *Escherichia coli:* interaction of the extracellular enzyme with human plasminogen and fibrinogen. *Int J Biochem Cell Biol* 2007; **39**: 1190-203.
183. Aguilera L, Ferreira E, Gimenez R et al. Secretion of the housekeeping protein glyceraldehyde-3-phosphate dehydrogenase by the LEE-encoded type III secretion system in enteropathogenic *Escherichia coli*. *Int J Biochem Cell Biol* 2012; **44**: 955-62.
184. Smith JL, Fratamico PM, Gunther NW. Extraintestinal pathogenic *Escherichia coli*. *Foodborne Pathog Dis* 2007; **4**: 134-63.
185. WHO. Antimicrobial resistance: global report on surveillance. *World Health Organization* 2014; **http://www.who.int/drugresistance/documents/surveillancereport/en/**.

186. Brumbaugh AR, Mobley HL. Preventing urinary tract infection: progress toward an effective *Escherichia coli* vaccine. *Expert Rev Vaccines* 2012; **11**: 663-76.
187. Ahmed T, Bhuiyan TR, Zaman K et al. Vaccines for preventing enterotoxigenic *Escherichia coli* (ETEC) diarrhoea. *Cochrane Database Syst Rev* 2013: CD009029.
188. Russo TA, Johnson JR. Extraintestinal isolates of *Escherichia coli*: identification and prospects for vaccine development. *Expert Rev Vaccines* 2006; **5**: 45-54.
189. O'Hanley P, Marcus R, Baek KH et al. Genetic conservation of hlyA determinants and serological conservation of HlyA: basis for developing a broadly cross-reactive subunit *Escherichia coli* alpha-hemolysin vaccine. *Infection and immunity* 1993; **61**: 1091-7.
190. O'Hanley P, Lalonde G, Ji G. Alpha-hemolysin contributes to the pathogenicity of piliated digalactoside-binding *Escherichia coli* in the kidney: efficacy of an alpha-hemolysin vaccine in preventing renal injury in the BALB/c mouse model of pyelonephritis. *Infection and immunity* 1991; **59**: 1153-61.
191. Pecha B, Low D, O'Hanley P. Gal-Gal pili vaccines prevent pyelonephritis by piliated *Escherichia coli* in a murine model. Single-component Gal-Gal pili vaccines prevent pyelonephritis by homologous and heterologous piliated *E. coli* strains. *J Clin Invest* 1989; **83**: 2102-8.
192. Schmidt MA, O'Hanley P, Lark D, Schoolnik GK. Synthetic peptides corresponding to protective epitopes of *Escherichia coli* digalactoside-binding pilin prevent infection in a murine pyelonephritis model. *Proc Natl Acad Sci U S A* 1988; **85**: 1247-51.
193. Roberts JA, Hardaway K, Kaack B et al. Prevention of pyelonephritis by immunization with P-fimbriae. *J Urol* 1984; **131**: 602-7.
194. Roberts JA, Kaack MB, Baskin G et al. Antibody responses and protection from pyelonephritis following vaccination with purified *Escherichia coli* PapDG protein. *J Urol* 2004; **171**: 1682-5.
195. Grischke EM, Ruttgers H. Treatment of bacterial infections of the female urinary tract by immunization of the patients. *Urol Int* 1987; **42**: 338-41.
196. Bauer HW, Rahlfs VW, Lauener PA, Blessmann GS. Prevention of recurrent urinary tract infections with immuno-active *E. coli* fractions: a meta-analysis of five placebo-controlled double-blind studies. *Int J Antimicrob Agents* 2002; **19**: 451-6.
197. Steffen R, Castelli F, Dieter Nothdurft H et al. Vaccination against enterotoxigenic *Escherichia coli*, a cause of travelers' diarrhea. *J Travel Med* 2005; **12**: 102-7.
198. Walker RI, Steele D, Aguado T, Ad Hoc ETEC. Analysis of strategies to successfully vaccinate infants in developing countries against enterotoxigenic *E. coli* (ETEC) disease. *Vaccine* 2007; **25**: 2545-66.
199. Clemens JD, Sack DA, Harris JR et al. Cross-protection by B subunit-whole cell cholera vaccine against diarrhea associated with heat-labile toxin-producing enterotoxigenic *Escherichia coli*: results of a large-scale field trial. *J Infect Dis* 1988; **158**: 372-7.
200. Svennerholm AM, Tobias J. Vaccines against enterotoxigenic *Escherichia coli*. *Expert Rev Vaccines* 2008; **7**: 795-804.
201. Quintana Flores VM, Campos de Souza Fernandes RC, Sousa de Macedo Z, Medina-Acosta E. Expression and purification of the recombinant enteropathogenic *Escherichia coli* vaccine candidates BfpA and EspB. *Protein Expr Purif* 2002; **25**: 16-22.
202. de Souza Campos Fernandes RC, Quintana Flores VM, Sousa de Macedo Z, Medina-Acosta E. Coproantibodies to the enteropathogenic *Escherichia coli* vaccine candidates BfpA and EspB in breastfed and artificially fed children. *Vaccine* 2003; **21**: 1725-31.
203. Donnenberg MS, Tacket CO, Losonsky G et al. Effect of prior experimental human enteropathogenic *Escherichia coli* infection on illness following homologous and heterologous rechallenge. *Infection and immunity* 1998; **66**: 52-8.

204. Gu J, Liu Y, Yu S et al. Enterohemorrhagic *Escherichia coli* trivalent recombinant vaccine containing EspA, intimin and Stx2 induces strong humoral immune response and confers protection in mice. *Microbes Infect* 2009; **11**: 835-41.
205. Wan CS, Zhou Y, Yu Y et al. B-cell epitope KT-12 of enterohemorrhagic *Escherichia coli* O157:H7: a novel peptide vaccine candidate. *Microbiol Immunol* 2011; **55**: 247-53.
206. Cai K, Gao X, Li T et al. Enhanced immunogenicity of a novel Stx2Am-Stx1B fusion protein in a mice model of enterohemorrhagic *Escherichia coli* O157:H7 infection. *Vaccine* 2011; **29**: 946-52.
207. Cai K, Gao X, Li T et al. Intragastric immunization of mice with enterohemorrhagic *Escherichia coli* O157:H7 bacterial ghosts reduces mortality and shedding and induces a Th2-type dominated mixed immune response. *Can J Microbiol* 2010; **56**: 389-98.
208. Gu J, Ning Y, Wang H et al. Vaccination of attenuated EIS-producing Salmonella induces protective immunity against enterohemorrhagic *Escherichia coli* in mice. *Vaccine* 2011; **29**: 7395-403.
209. Riquelme-Neira R, Rivera A, Saez D et al. Vaccination with DNA Encoding Truncated Enterohemorrhagic *Escherichia coli* (EHEC) Factor for Adherence-1 Gene (efa-1') Confers Protective Immunity to Mice Infected with *E. coli* O157:H7. *Front Cell Infect Microbiol* 2015; **5**: 104.
210. Bentancor LV, Bilen M, Brando RJ et al. A DNA vaccine encoding the enterohemorragic *Escherichia coli* Shiga-like toxin 2 A2 and B subunits confers protective immunity to Shiga toxin challenge in the murine model. *Clin Vaccine Immunol* 2009; **16**: 712-8.
211. Fraser CM, Rappuoli R. Application of microbial genomic science to advanced therapeutics. *Annu Rev Med* 2005; **56**: 459-74.
212. Andre FE. Vaccinology: past achievements, present roadblocks and future promises. *Vaccine* 2003; **21**: 593-5.
213. Fleischmann RD, Adams MD, White O et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; **269**: 496-512.
214. Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 2001; **19**: 2688-91.
215. Rappuoli R. Reverse vaccinology. *Curr Opin Microbiol* 2000; **3**: 445-50.
216. Rappuoli R, Bottomley MJ, D'Oro U et al. Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *J Exp Med* 2016; **213**: 469-81.
217. Pizza M, Scarlato V, Masignani V et al. Identification of vaccine candidates against serogroup B *meningococcus* by whole-genome sequencing. *Science* 2000; **287**: 1816-20.
218. Tettelin H, Saunders NJ, Heidelberg J et al. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 2000; **287**: 1809-15.
219. Maione D, Margarit I, Rinaudo CD et al. Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* 2005; **309**: 148-50.
220. Thorpe C, Edwards L, Snelgrove R et al. Discovery of a vaccine antigen that protects mice from *Chlamydia pneumoniae* infection. *Vaccine* 2007; **25**: 2252-60.
221. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 2012; **13**: 227-32.
222. Medini D, Serruto D, Parkhill J et al. Microbiology in the post-genomic era. *Nature reviews Microbiology* 2008; **6**: 419-30.
223. Larsson O, Nadon R. Gene expression - time to change point of view? *Biotechnol Genet Eng Rev* 2008; **25**: 77-92.
224. Sanchez C, Lachaize C, Janody F et al. Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database. *Nucleic Acids Res* 1999; **27**: 89-94.

225. Sargeant TJ, Marti M, Caler E et al. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* 2006; **7**: R12.
226. Cullen PA, Xu X, Matsunaga J et al. Surfaceome of *Leptospira* spp. *Infection and immunity* 2005; **73**: 4853-63.
227. Solis N, Cordwell SJ. Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics* 2011; **11**: 3169-89.
228. Kawasaki S, Moriguchi R, Sekiya K et al. The cell envelope structure of the lipopolysaccharide-lacking gram-negative bacterium *Sphingomonas paucimobilis*. *Journal of bacteriology* 1994; **176**: 284-90.
229. Cottingham K. Shaving proteins off the plasma membrane. *J Proteome Res* 2006; **5**: 738.
230. Abath FG, Almeida AM, Ferreira LC. Identification of surface-exposed *Yersinia pestis* proteins by radio-iodination and biotinylation. *J Med Microbiol* 1992; **37**: 420-4.
231. Maddalo G, Stenberg-Bruzell F, Gotzke H et al. Systematic analysis of native membrane protein complexes in *Escherichia coli*. *J Proteome Res* 2011; **10**: 1848-59.
232. Santoni V, Molloy M, Rabilloud T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* 2000; **21**: 1054-70.
233. Olaya-Abril A, Jimenez-Munguia I, Gomez-Gascon L, Rodriguez-Ortega MJ. Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. *J Proteomics* 2014; **97**: 164-76.
234. Walters MS, Mobley HL. Identification of uropathogenic *Escherichia coli* surface proteins by shotgun proteomics. *J Microbiol Methods* 2009; **78**: 131-5.
235. Rybak JN, Scheurer SB, Neri D, Elia G. Purification of biotinylated proteins on streptavidin resin: a protocol for quantitative elution. *Proteomics* 2004; **4**: 2296-9.
236. Poetsch A, Wolters D. Bacterial membrane proteomics. *Proteomics* 2008; **8**: 4100-22.

CHAPTER II: OBJECTIVES

OBJECTIVES

The aim of my research work was to identify some novel adhesins in EHEC O157:H7 in a strategy for preventing intestinal colonisation. Rather than classical reverse vaccinology approach solely based on genomic analysis aiming at identifying potential antigen candidates based on the prediction of their subcellular location at the bacterial cell surface, the identification of candidates was here primarily based on subproteomic analysis. To investigate the cell surface proteome, an original approach based on the use of biotinylation reagents of different sizes was first developed. Following the secretome concept, which considers globally both the protein secretion systems and the cognate secreted proteins, subproteomic analysis of the prototypical EHEC, namely *E. coli* O157:H7 EDL933, was further performed considering the protesurfaceome (cell-surface proteome), exoproteome (extracellular proteome) and proteovesiculome (membrane-vesicle proteome). In parallel, the complement of secretome participating to bacterial adhesion was reviewed following a literature survey. Focusing on the bacterial cell surface, proteosurfaceome analyses were also carried out to identify some targets of interest potentially acting as colonisation factors towards some extracellular matrix (ECM) molecules. Finally, some of these surface proteins were characterised with respect to their cell envelope subcellular localization, interaction with the ECM and adhesion to intestinal epithelial cells.

CHAPTER III: DIFFERENTIAL BIOTIN LABELLING OF THE CELL ENVELOPE PROTEINS IN LIPOPOLYSACCHARIDIC DIDERM BACTERIA: EXPLORING THE PROTEOSURFACEOME OF *ESCHERICHIA COLI* USING SULFO-NHS-SS-BIOTIN AND SULFO-NHS-PEG4-BISMANNOSE-SS-BIOTIN

Differential Biotin Labelling of the Cell Envelope Proteins in Lipopolysaccharidic Diderm Bacteria: Exploring the Proteosurfaceome of *Escherichia coli* using sulfo-NHS-SS-biotin and sulfo-NHS-PEG4-bismannose-SS-biotin

Ricardo MONTEIRO [1,2], Ingrid CHAFSEY [1], Sabine LEROY [1], Christophe CHAMBON [3], Michel HÉBRAUD [1,3], Valérie LIVRELLI [4], Mariagrazia PIZZA [2], Alfredo PEZZICOLI [2], Mickaël DESVAUX [1,*]

Submitted Journal of Proteomics, October 2017

[1] Université Clermont Auvergne, INRA, UMR454 MEDiS, F-63000 Clermont-Ferrand, France.

[2] GSK, Via Fiorentina 1, 53100, Siena, Italy.

[3] INRA, Plate-Forme d'Exploration du Métabolisme, F-63122 Saint-Genès Champanelle, France.

[4] Centre de Recherche en Nutrition Humaine Auvergne, UMR UCA INSERM U1071, USC-INRA 2018, Clermont Université - Université d'Auvergne, Faculté de Pharmacie, CHU Clermont-Ferrand, Service Bactériologie Mycologie Parasitologie, Clermont-Ferrand, France

*Corresponding author: Dr Hab Mickaël DESVAUX-LENÔTRE, INRA (Institut National de la Recherche Agronomique) Auvergne-Rhône-Alpes, Centre de Recherche de Clermont-Ferrand, UMR454 MEDiS (Microbiologie, Environnement Digestif, Santé), Site de Theix, F-63122 Saint-Genès Champanelle, France. Tel.: +33 (0)4 73 62 47 23, Fax: +33 (0)4 73 62 45 81, E-mail: mickael.desvaux@inra.fr.

ABSTRACT

Surface proteins are the major factor for the interaction between bacteria and its environment, playing an important role in infection, colonisation, virulence and adaptation. However, the study of surface proteins has proven difficult mainly due to their hydrophobicity and/or relatively low abundance compared with cytoplasmic proteins. To overcome these issues new proteomic strategies have been developed, such as cell-surface protein labelling using biotinylation reagents. Sulfo-NHS-SS-biotin is the most commonly used reagent to investigate the proteins expressed at the cell surface of various organisms but its use in lipopolysaccharidic diderm bacteria (archetypical Gram-negative bacteria) remains limited to a handful of species. While generally pass over in silence, some periplasmic proteins, but also some inner membrane lipoproteins (IMLs), integral membrane proteins (IMPs) and cytoplasmic proteins (cytoproteins) are systematically identified following this approach. To limit cell lysis and diffusion of the sulfo-NHS-SS-biotin through the outer membrane, biotin labelling was tested over short incubation times and proved to be as efficient for 1 min at room temperature. To further limit labelling of protein located below the outer membrane, the use of high-molecular weight sulfo-NHS-PEG4-bismannose-SS-biotin appeared to recover differentially cell-envelope proteins compared to low-molecular weight sulfo-NHS-SS-biotin. Actually, the sulfo-NHS-SS-biotin recovers at a higher extent the proteins completely or partly exposed in the periplasm than sulfo-NHS-PEG4-bismannose-SS-biotin, namely periplasmic proteins, outer membrane lipoproteins (OMLs), IMLs and IMPs. These results highlight that protein labelling using biotinylation reagents of different sizes provides a sophisticated and accurate way to differentially explore the cell envelope proteome of lipopolysaccharidic diderm bacteria.

INTRODUCTION

Protein secretion plays a key role in the interactions of a bacterial cell with its environment [1-3]. Extracytoplasmic proteins have a wealth of physiological functions, ranging from degradative enzymes, colonisation to virulence factors, which are relevant to the role and place of the microorganism in its ecosystem, *e.g.* as saprophyte, symbiont or pathogen [4]. In lipopolysaccharidic diderm bacteria (archetypal Gram-negative bacteria) and respective to the Gene Ontology (GO) for cellular components, exported proteins (transported across the cytoplasmic membrane) can sublocalise at the inner membrane (IM) (GO: 0005886), in the periplasm (PE) (GO: 0042597) or anchored on the periplasmic side of the outer membrane (OM) (GO: 0036406), whereas secreted proteins (further transported across the OM) can localise at the outer membrane (GO: 0019867), extracellular milieu (EC) (GO: 0005576) or even within the cytosol of a host cell [5]. Respective to the cell envelope of LPS-diderm bacteria, surface proteins are localised at the OM and can be further discriminated into these intrinsic to the OM, *i.e.* integral to the OM (GO: 0045203) or anchored to OM (GO: 0036405) and exposed on the external side, or extrinsic to the OM on the external side (GO: 0031242).

To date, 9 protein secretion systems (numbered from Type I to Type IX) have been uncovered in LPS-diderm bacteria, allowing protein transport across the cell envelope, that is from the interior to the exterior of the bacterial cell [5, 6]. According to proteogenomic analyses, secreted proteins exposed at the cell surface can represent about one-third of the entire encoded bacterial proteins [7] but experimental access to the cell-surface proteome (proteosurfaceome) has been proven difficult. In fact, disclosing this subproteome in LPS-diderm bacteria by classical proteomic approaches is impaired by numerous factors, such as the hydrophobicity of outer membrane proteins (OMPs) or the low abundance of cell surface proteins [8, 9]. To circumvent these limitations, alternative proteomic

methods for the analysis of the bacterial surface proteins have been developed over the last decade [10]. Due to the issues of OMPs solubility and/or surface protein abundance [11], off-gel proteomics now clearly exceed in-gel approaches for protein separation when investigating the proteosurfaceome [10]. Basically, techniques for producing surface-enriched protein preparations can be discriminated into (i) subcellular fractionation [12], (ii) cell surface shaving [13], and (iii) cell surface labelling [13, 14]. Historically, subcellular fractionation was the first to be developed to isolate membrane proteins but low reproducibility, low yields and/or high contamination levels by cytoplasmic proteins (cytoproteins) were generally reached. A major advance aroused with the cell-surface shaving by proteolysis, which allowed circumventing the problem related to the OMPs extraction by simply cleaving the hydrophilic protein regions exposed at the cell surface, generally using trypsin. However, the release of cytoproteins generally occurs because of cell lysis during the incubation with the protease. More recently, cell surface labelling approaches were developed, especially when using biotinylation [15].

Based on its successful use in various eukaryotic cells, the first attempts of surface-protein biotinylation in LPS-diderm bacteria were performed in *Escherichia coli* [16], *Helibacter pylori* [17], *Leptospira* spp. [14], *Rickettsia parkeri* [18] and *Yersinia pestis* [13, 19] with sulfo-NHS-LC-biotin (sulfo-succinimidyl biotin-amidohexanoate). The use of this water-soluble biotinylation reagent, primarily cross-linking with exposed primary amine groups in proteins [20], appeared promising to decipher the bacterial proteosurfaceome but the strong affinity between biotin and avidin led to poor recovery of the biotinylated proteins from the affinity column. To facilitate the extremely difficult elution of the labelled proteins from the avidin support, introduction of a disulphide bridge was introduced in the linker region of the biotinylation reagent to allow cleavage under reducing conditions [21, 22]. In LPS-diderm bacteria, the use of sulfo-NHS-SS-biotin remains limited to a handful of

species, namely *Ehrlichia chaffeensis* [23], *Anaplasma phagocytophilum* [24], uropathogenic *E. coli* [13], *Neorickettsia sennetsu* [25], *Shewanella oneidensis* [26], *Rickettsia* spp. [27-29], and *Bacteroides fragilis* [30]. While considered as plasma membrane impermeable in eukaryotic cells because of the negatively charged sulfo group, it appeared that some periplasmic, cytoplasmic and IM proteins were also identified by these approaches. Besides cell lysis that could occur during the procedure (especially at centrifugation and incubation stages), it is also known that hydrophilic molecules (including biotinylation reagents) can passively cross the OM through porins, with a size exclusion limit estimated at 600-800 Da [17, 31]. To investigate the effect of the molecular size of the biotinylation reagent on protein cell surface labelling in LPS-diderm bacteria, *E. coli* was here used as a model organism to characterise its proteosurfaceome by using a soluble, cleavable and high-molecular-weight biotin derivative, namely sulfo-NHS-PEG4-bismannose-SS-biotin (1472 Da), in comparison to the classical, soluble and cleavable sulfo-NHS-SS-biotin (607 Da).

MATERIAL AND METHODS

Bacterial growth conditions

E. coli K12 strain MG1655 was used for this study. One isolated colony was taken off from LB (Lysogeny Broth) agar plate, inoculated and grown overnight in LB liquid media at 37 °C under orbital shaking. After 1:100 dilution, bacterial cultures were grown in the same conditions until late exponential phase (0.8 $OD_{600\ nm}$).

Biotinylation of bacterial cell surface proteins and protein affinity purification

Bacterial cells were harvested by centrifugation (4000 *g*, 5 min, 4 °C) and washed twice with 10 mM PBS pH 8 (4000 *g*, 5 min, 4 °C). Pelleted cells were weighed and resuspended in 10 mM PBS pH 8. Biotin-protein labelling reaction was performed incubating bacterial cells with sulfo-NHS-SS-biotin (Sulfo-succinimidyl biotin-amidoethyl dithio-propionate; 606.69 Da; Thermo Scientific) or sulfo-NHS-PEG4-bismannose-SS-biotin (Sulfo-succinimidyl biotin-amidoethyl dithio-propionyl-amino-benzoyl-bismannose propylamino-tetraoxododecanyl; 1471.54 Da; Interchim) at 1 % (m/m), during 1, 5, 15 and 30 minutes with gentle agitation at room temperature. Excess of the biotinylation reagent was quenched by three washes with a solution of 10 mM PBS, pH 8 and 500 mM Glycine (4000 *g*, 5 min, 4 °C) and cells were resuspended in lysis buffer (10 mM PBS pH 8, 1 mM PMSF, 1 % v/v Triton X-100). As reaction control, cells were incubated with 10 mM PBS, pH 8 instead of biotin reagent and underwent the same procedure. Cell disruption was performed using Fast-prep (MP Biomedicals) with two steps of 20 s at 6 m/s and the cell debris from cell lysis discarded by centrifugation (20 000 *g*, 30 min, 4 °C). Labelled proteins in the supernatant were purified by affinity purification over column containing NeutrAvidin agarose resin (Pierce Thermo Scientific) following the manufacture instructions with some modifications. Briefly, columns were equilibrated with wash buffer (10 mM PBS, pH 8, 1% NP-

40) and the exactly same volume of each protein samples as well the control was load and kept in contact with the resin at room temperature for 15 min. Unlabelled proteins were washed away by 10 column volumes with wash buffer. Finally, labelled proteins were eluted (2 % SDS, 20 % glycerol, 62.5 mM Tris-HCl, 50 mM DTT, 5 % β-mercaptoethanol).

Mass spectrometry analysis

All eluted samples were loaded in a SDS-PAGE gel and concentrated in a single band. Excised bands were then washed in 25 mM ammonium bicarbonate with 5 % acetonitrile for 30 min and twice in 25 mM ammonium bicarbonate with 50 % acetonitrile for 30 min. Reduction and alkylation reactions were performed with 10 mM DTT and 55 mM iodoacetamide solutions, respectively, before the bands were dehydrated with 100 % acetonitrile. The samples were hydrolysed overnight with 600 ng of trypsin (Promega) and the peptides extracted with 100 % acetonitrile before being concentrated (SpeedVac) and resuspended in a same voulme.

Hydrolysed samples were analysed by nanoLC-MS/MS using an Ultimate 3000 system (Dionex Thermo Fisher) coupled to an Impact II QTOF (Bruker Daltonics). After desalting on a C18 pre-column (300 µm, 5 mm), peptides were separated on an analytical C18 nanocolumn (75 µm, 15 cm) using a 50 min gradient from 10 % to 40 % solvent A (80/20 acetonitrile/H_2O, 0.5 % Formic acid) in solvent A (100 % H_2O, 0.1 % Formic acid). The eluate was electrosprayed in the impact II using nanoCaptiveSpray source (Bruker Daltonics). Mass spectrometer was used in CID (collision-induced dissociation) mode to acquire a maximum number of MS/MS in 3 s after the full MS scans. Each condition was performed in triplicates with two runs per sample.

Bioinformatic analysis

The proteins identification was performed from *E. coli* K12 MG1655 database (NCBI Accession U00096, Version U00096.3) using using ProteinScape v1.1 (Brucker). Peptides were validated for a

Mascot percolator score permitting to obtain a false discovery rate (FDR) below at 1 % and proteins were identified when a single high-quality tandem MS spectrum of a peptide matched significantly with the database and validated for a Mascot score higher than 17. Relative quantitation of proteins was performed following the workflow provided by Progenesis QI for proteomics (Nonlinear dynamics, Waters). Differential protein quantification was considered significant for fold abundance greater than 1.5 and *p*-value lower than 0.05. For prediction of protein subcellular localisation (SCL) in *E. coli* K12 MG1655, the validated proteins were compared with the combined database generated by Díaz-Mejía *et al*. [32]. The enrichment of gene ontology (GO) for molecular function was performed using STRAP (Software for Researching Annotations of Proteins) [33], which retrieved overrepresented GO terms related with molecular function based on UniprotKB and EBI QuickGO. Identified proteins predicted as located at the cytoplasm were filtered for moonlight activity using a moonlight proteins database (http://www.moonlightingproteins.org).

RESULTS

Efficient labelling of cell envelope proteins with sulfo-NHS-SS-biotin over short incubation times

To date, all studies investigating the bacterial proteosurfaceome use the sulfo-NHS-SS-biotin with an incubation time varying from about 30 min to 2h at 0 or 4°C [13, 22-28, 30, 34, 35]. To limit thermic shock and considering the reagent is extremely reactive at room temperature, a time scale of 1, 5, 15 and 30 min of incubation with sulfo-NHS-SS-biotin probe were tested prior to protein identification and relative protein quantitation. At the different incubation times, a total of 199, 116, 121, and 105 different proteins were identified, respectively (Table 1S, supplementary data). Based on the combined database generated by Díaz-Mejía *et al*. for SCL of proteins in *E. coli* MG1655 [28], none of the proteins here identified were predicted as located in the extracellular milieu (GO: 0005576) but as located either within the cell envelope (GO: 0030313) or in the cytoplasm (GO: 0005737). Ranging from 1 to 30 min incubation times, the percentages of proteins identified for each predicted SCL were quite close (Figure 1A). As expected, the majority of the proteins were predicted within the cell envelope (63.8-69.8 %), that is 8.6-10.6 % at the OM (GO: 0036406), 26.1-30.2 % in the periplasm (GO: 0042597), and 27.1-32.2 % at IM (GO: 0005886). Nonetheless, the highest percentage of proteins predicted as located at the OM (i.e. 10.6 %) was observed with 1 min incubation time; this is also with this incubation time that the percentages of proteins predicted in the periplasm and at the IM were the lowest, i.e. 26.1 and 27.1 %, respectively. In order to directly and quantitatively compare the proteins identified in these different biotinylation conditions, statistical analysis of the protein abundance at the four incubation times revealed no significant difference (supplementary material Table 1S). Altogether, this indicates that labelling of cell-envelope proteins with sulfo-NHS-SS-biotin was as efficient with

incubation time as short as 1 min at room temperature. However, even over short incubation times, it appears that some periplasmic and IM proteins were still labelled by sulfo-NHS-SS-biotin, suggesting it could cross the OM and reach other subcellular compartments.

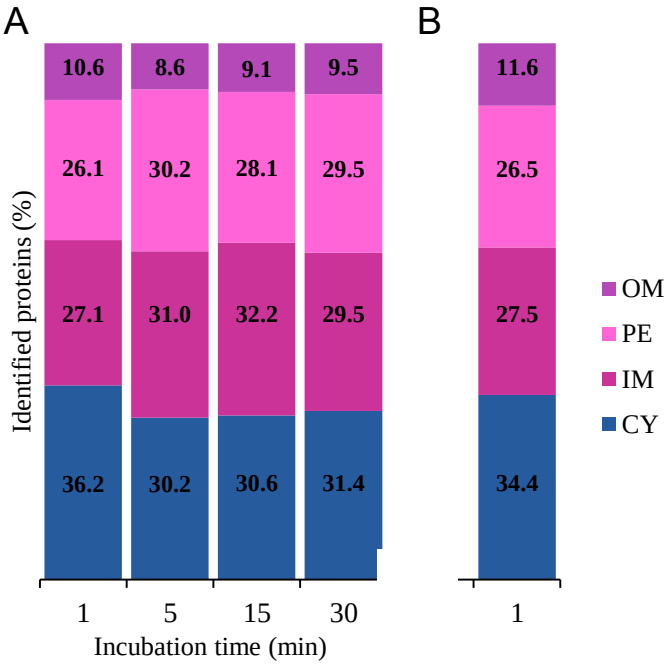


Figure 1: Effect of the incubation time (A) and molecular weight of the biotinylation reagents (B) on the percentage of proteins identified respective to their predicted SCL in *E. coli*. A: Bacterial cells were incubated with sulfo-NHS-SS-biotin for 1, 5, 15 and 30 min respectively. B: Bacterial cells were incubated for 1 min with sulfo-NHS-PEG4-bismannose-SS-biotin. Pink shades are related to proteins predicted as located within the cell envelope (GO: 0030313), *i.e.* either at the outer membrane (OM; GO: 0036406), in the periplasm (PE; GO: 0042597), or at the inner membrane (IM; GO: 0005886). Blue colour stands for proteins predicted as sublacted in the cytoplasm (CY; GO: 0005737).

Differential cell-envelope protein recovery with high-molecular weight sulfo-NHS-PEG4-bismannose-SS-biotin compared to low-molecular weight sulfo-NHS-SS-biotin

Considering the size limit for outer membrane porins, the use of a biotinylation reagent with a molecular weight much higher than 800 Da was considered, namely sulfo-NHS-PEG4-bismannose-SS-

biotin. Following the same workflow with a 1 min incubation time with sulfo-NHS-SS-biotin, proteomic analysis using sulfo-NHS-PEG4-bismannose-SS-biotin was performed. The majority of the identified proteins were predicted as localised within the cell envelope, namely 65.6% (Figure 1B); except for the percentage of proteins predicted as located at the OM, which was higher when using sulfo-NHS-PEG4-bismannose-SS-biotin than sulfo-NHS-SS-biotin, the percentage of proteins recovered by these biotinylation reagents and predicted as located in the periplasm and the OM were quite similar, namely 26.5 and 27.5 % with sulfo-NHS-PEG4-bismannose-SS-biotin.

Regarding the molecular functions, they were related to binding (GO: 0005488) and catalytic activity (GO: 0003824) for the majority of the identified proteins (78 %) (Figure 2); about 3 % and 5 % of the proteins were related to molecular transducer and structural activities, respectively, whereas the remaining 14 % cover several other individual functions, *i.e.* representing each less than 1 % (Figure 2A and supplementary material Table 1S). Discriminating identified proteins between those with a significantly and non-significantly different abundance when using sulfo-NHS-PEG4-bismannose-SS-biotin versus sulfo-NHS-SS-biotin (Figure 2BC), the proportion of identified proteins with molecular functions related to binding activity remain similar (42 %). For the catalytic activity, however, it reached 37 % for the identified proteins with no significant differences in abundance for the two reagents and decrease to 18 % for the identified proteins with significant differences when using sulfo-NHS-PEG4-bismannose-SS-biotin versus sulfo-NHS-SS-biotin. This suggests some protein subcategories could be under- or over-represented when using one of the two reagents.

In order to evaluate the differential abundance of proteins recovered with the high-molecular weight sulfo-NHS-PEG4-bismannose-SS-biotin versus the classical sulfo-NHS-SS-biotin, a label-free quantitative proteomic analysis was performed. Spectra profiles from hydrolysed peptides were analysed and relatively quantified resulting in 189 proteins validated for statistical comparative analysis

(supplementary material Table 1S). This revealed that ~34 % of the total quantified proteins (65 out of 189 identified proteins) presented significantly different abundance between the two biotinylation reagents. 53 out of 65 proteins were lower recovered and/or absent when using sulfo-NHS-PEG4-bismannose-SS-biotin and 12 were higher abundant and/or only recovered with this reagent.

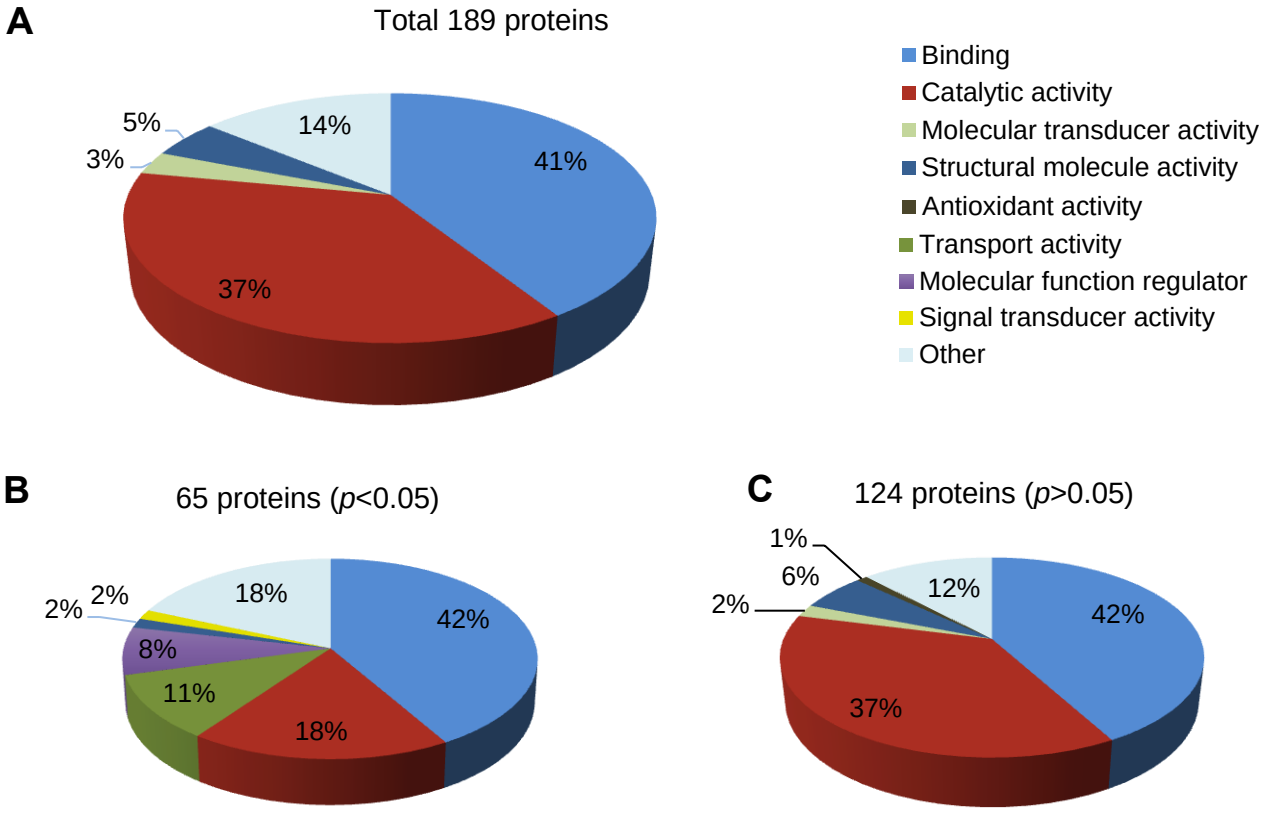


Figure 2: Distribution of the molecular function terms according to the gene ontology (GO) among the proteins identified following biotinylation with sulfo-NHS-PEG4-bismannose-SS-biotin and/or sulfo-NHS-SS-biotin. A: Proportions of the molecular function terms for the totality of the 189 proteins identified whenever with sulfo-NHS-PEG4-bismannose-SS-biotin or sulfo-NHS-SS-biotin. B: Proportions of the molecular function terms for the 65 proteins with statistically significant relative abundances (p<0.05) when using sulfo-NHS-PEG4-bismannose-SS-biotin compared to sulfo-NHS-SS-

biotin. C: Proportions of the molecular function terms for the 124 proteins with no significant statistical difference (p>0.05) in their relative abundance when using one or the other biotinylation reagent.

The sulfo-NHS-SS-biotin recovers at a higher extent the proteins completely or partly exposed in the periplasm than sulfo-NHS-PEG4-bismannose-SS-biotin

Considering the identified proteins predicted as localised at the cell envelope (Table 1), different subcategories of proteins could be discriminated respective to their SCL. At the OM, several (i) subunits of cell surface organelles, namely the flagella, (ii) outer membrane proteins (OMPs) anchored via a β-barrel, *i.e.* integral to the OM (GO: 0045203), and (iii) outer membrane lipoproteins (OMLs), *i.e.* anchored on the internal side of the OM (GO: 0036406), could be here identified. With sulfo-NHS-PEG4-bismannose-SS-biotin, the flagellin FliC and flagellar hook FlgE were recovered at significantly higher levels than with sulfo-NHS-SS-biotin. These flagella subunits are found at the cell surface (GO: 000986) and associate to form the flagella supramolecular complex, itself anchored to the OM via the flagellum secretion and assembly machinery, namely the Type III subtype b secretion system (T3bSS) [36, 37]. Some key OMPs were here identified (Table 1), such as the exit duct TolC [38], the porin OmpA [39], the autoaggregative factor Ag43 [40], or the β-barrel assembly machinery factor A (BamA) [41]. While no significant difference could be observed for the recovery of most of these OMPs when using the two biotinylation reagents (Table 1), the maltoporin LamB could only be identified when using the sulfo-NHS-PEG4-bismannose-SS-biotin. Regarding OMLs, some were recovered at significantly higher levels with the low molecular weight biotinylation reagent sulfo-NHS-SS-biotin. Interestingly, those characterised OMLs were systematically anchored to the internal side of the OM (GO: 0036406), *e.g.* Pal, which links to the peptidoglycan through its periplasmic domain [42], LpoA, which spans the periplasm to stimulate the peptidoglycan synthase PBP1A [43], or BamD,

which associates on the periplasmic side to BamA together with the OMLs BamBCE to form the Bam complex [41]. On the contrary, some of the characterised OMLs with no significant difference in the level of recovery using one or the other biotinylation reagents are known to protrude on the external side of the OM, *e.g.* Lpp [44], Pcp [45], or BamC [46].

While numerous periplasmic proteins could be here identified with both biotinylation reagents, 68 % of them (34 out of 50) were recovered at significantly lower levels with sulfo-NHS-PEG4-bismannose-SS-biotin and some of them could only be identified with the sulfo-NHS-SS-biotin, *i.e.* YbiS, GltI and HisJ (Table 1). Besides, none of the identified periplasmic proteins were found at significantly higher abundance using sulfo-NHS-PEG4-bismannose-SS-biotin. Together with the differential recovery of some OMLs, this indicates the high molecular weight biotinylation reagent recovers at a lower extent some of the proteins that are completely or partly exposed in the periplasm. Considering the OMLs that are not surface exposed, the periplasmic proteins, the inner membrane proteins (IMPs) and inner membrane lipoproteins (IMLs), 47 % of them were only recovered or at significantly higher levels with the low molecular weight sulfo-NHS-SS-biotin (Table 1). Altogether, this indicates the sulfo-NHS-SS-biotin is more prone to recover proteins located below the OM than sulfo-NHS-PEG4-bismannose-SS-biotin.

Interestingly, 23 out of the 65 proteins primarily predicted as located in the cytoplasm were predicted as moonlighting proteins, *i.e.* as extrinsic to the OM on the external side (GO: 0031242). As numerous glycolytic enzymes [47], glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is involved in binding to plasminogen and fibrinogen when exposed on the cell surface of *E. coli* [48, 49]. In this species, enolase is well-known to be secreted, which is regulated by an automodification process [50]. In *E. coli*, the elongation factor Tu (EF-Tu) is also known to be cell surface exposed and involved in bacterial aggregation [51].

DISCUSSION

For the record, the use of sulfo-NHS-SS-biotin in bacteria was for the first time performed in a bacterium with a Gram-positive cell envelope (i.e. parietal monoderm bacteria), namely *S. aureus* [52], before being tested in LPS-diderm bacteria, first in *E. chaffeensis*, where it proved to be much more suitable and appropriate for proteomics than sulfo-NHS-LC-biotin [23]. While successfully applied for investigating the proteosurfaceome of other LPS-diderm bacteria, and contrary to investigations in parietal monoderm bacteria [34, 35, 52-54], two-dimensional gel electrophoresis or shotgun proteomic analyses systematically identified some periplasmic proteins and OMLs, but also some IMLs and IMPs [13, 25-30]. While generally pass over in silence, their identification suggests the biotinylation reagent is not impermeable but crosses the OM through porins [55, 56], and this could further be increased over incubation at low temperatures [57]; indeed, the biophysical properties of the OM, including its permeability, are different from the (cyto)plasmic membrane, for which the biotinylation reagents are rendered impermeable due to the sodium sulfoxide group on the succinimidyl ring [22, 58]. In order to limit the possible diffusion through the OM and/or cell lysis that could occur in the course of long incubation times (> 30 min) with sulfo-NHS-SS-biotin, incubation at room temperature as short as 1 min was here proven to be as efficient. However, some proteins predicted as located below the OM could still be identified.

Considering that porins are responsible for the OM permeability, with size exclusion limit estimated at < 800 Da [31], we tested a biotinylation reagent with high molecular weight, i.e. sulfo-NHS-PEG4-bismannose-SS-biotin. As expected, the cell-envelope proteins recovered upon labelling with this reagent was different from the classical low molecular weight sulfo-NHS-SS-biotin. A significant number of periplasmic proteins, OMLs, IMLs and IMPs were either not or recovered at lower levels

with sulfo-NHS-PEG4-bismannose-SS-biotin. These results highlight that protein labelling using biotinylation reagents of different sizes provides a sophisticated and elegant way to differentially explore the cell envelope proteome of LPS-diderm bacteria (Figure 3). Depending on the scientific questions ask, biotinylation reagents of small size can be used to explore deeply the bacterial cell envelope and label proteins on both side of the OM, periplasm and periplasmic side of IM, whereas larger biotinylation reagents can be used to mainly label proteins on the external side of the OM. Nonetheless, some periplasmic proteins, OMLs, IMLs and IMPs were still recovered with a biotinylation reagent as large as sulfo-NHS-PEG4-bismannose-SS-biotin, i.e. 1472 Da. This suggests that despite the maximum exclusion size limit, such a reagent would still pass through porins. Actually, at the time of these estimations still in effect, the molecular structure of OM porins were not known. Besides its molecular weight, the structure of a given molecule is certainly a much more important aspect to comprehend how its fit a porin channel with respect to its shape, surface charges and/or diameter [55, 59]. The molecular structure of both the labelling reagent and the targeted protein can be affected by the steric bulk, which in turn could influence the level of biotinylation of given proteins, like some OMPs [16]. So far the range of water-soluble biotinylation reagents commercially available and harbouring a disulphide bond in their linker region is quite limited to sulfo-NHS-SS-biotin and sulfo-NHS-PEG4-bismannose-SS-biotin. By stressing this need, the present investigation should hopefully trigger interest in developing a much broader range of such biotinylation reagents of various sizes and molecular structures, with different permeability to membranes, especially some molecules truly impermeable to the bacterial OM.

Several cytoproteins, i.e. proteins primarily predicted as localised in the cytoplasm, were predicted as moonlighting proteins. Among them, GAPDH and EF-Tu were indeed demonstrated as displaying additional function unrelated to their primary function when exposed at the cell surface of *E. coli* [44,

48], whereas enolase was at least reported as secreted by non-classical secretion in this species [50]. Surprisingly enough, this stresses that so far very few investigations have been dedicated to the characterisation of putative moonlighting proteins in a model organism like *E. coli* compared to some other bacterial species like *Streptococcus* [60]. In addition to enolase reported as binding plasmin(ogen) in numerous bacterial species [48, 49], ribosomal proteins are often found to be exposed on the bacterial surface and could display high immunogenicity to humans [61], phosphoglycerate kinase was demonstrated to bind plasmin(ogen) and actin in *Streptococcus* [62, 63], or triophosphate isomerase was reported to bind to intestinal epithelial cells in *Lactobacillus* [64]. In *E. coli*, these possible moonlighting functions still remain to be demonstrated, which opens the way to intense investigations in the field. While generically referred as transported by non-classical secretion (that is by unknown protein secretion systems), such proteins can also be transported by piggybacking via known secretion systems, via holins or mechano-sensitive channels (Msc), by outer membrane vesicles, or released upon controlled bacterial cell lysis, e.g. allolysis [65].

The outstanding and comprehensive database generated by Díaz-Mejía *et al*. [32] for protein subcellular localisation in *E. coli* K12 supported by the scientific literature and/or a wide range of prediction tools could help and backup the interpretation of our subproteomic investigation. Respective to less investigated bacterial species like *Neorickettsia sennetsu* [25] or *Shewanella oneidensis* [26], the lack of such information can limit the interpretation of proteomic results. Like it was developed for parietal monoderm bacteria [66], a comprehensive, generic and rational strategy based on the secretome, that is the biology of protein secretion in LPS-diderm bacteria, is still missing. The development of such dedicated analytical pipeline integrating exhaustive bioinformatic analyses and considering the protein export and secretion systems, as well as cognate post-translocational maturation

pathways, is crucial for insightful interpretation of results in LPS-diderm bacteria, especially the exoproteome and proteosurfaceome of important LPS-diderm bacterial pathogens.

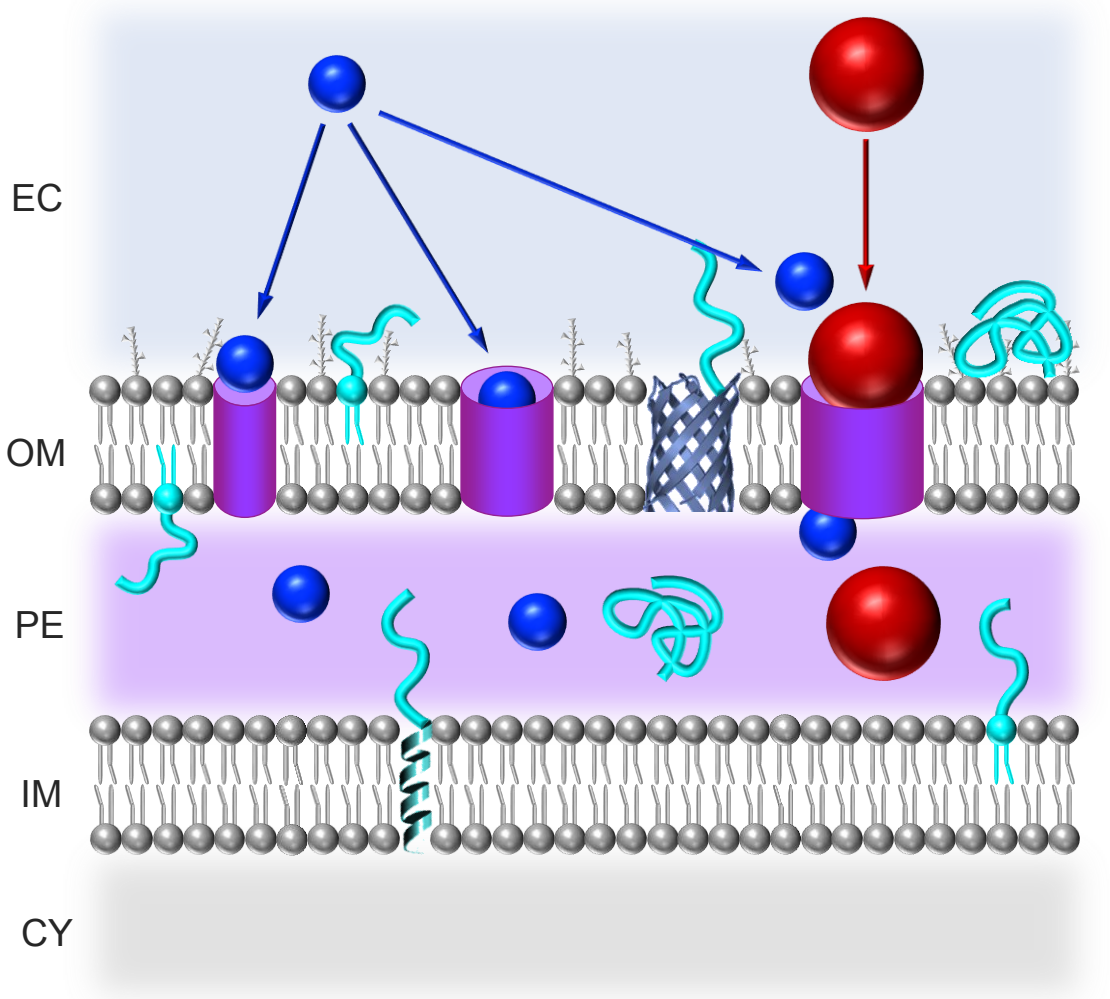


Figure 3: Schematic overview of differential cell-envelope protein labelling by biotinylation reagents in LPS-diderm bacteria. Biotinylation reagents can label the parts of proteins exposed on the external side of the OM, namely some surface organelles, OMPs, or protein associated to the cell surface (e.g. some cytoproteins), as well as some cell surface exposed OMLs. Depending on the molecular weight and molecular structure of the biotinylation reagents (i.e. high and low molecular weight biotinylation reagents as depicted in red and blue spheres, respectively), they can penetrate more or less efficiently into the bacterial cell envelope via porins and further label some periplasmic proteins, OMLs, IMLs and IMPs, whose parts are exposed in the periplasmic space.

Table 1. Proteins identified following biotinylation with sulfo-NHS-SS-biotin and sulfo-NHS-PEG4-bismannose-SS-biotin and predicted as localised in the cell envelope.

Uniprot ID	Name	Annotation	SCL [a] prediction [a]	Fold change [b]	*p*-value
Subunits of cell surface organelles					
P04949	FliC	Flagellin	OM	+6.1	5.82×10^{-3}
P75937	FlgE	Flagellar hook protein	OM	+5.1	1.79×10^{-3}
Outer membrane proteins (OMPs)					
P02943	LamB	Maltoporin	OM	+	5.60×10^{-6}
P02930	TolC	Outer membrane protein TolC	OM	=	$>5.00 \times 10^{-2}$
P0A910	OmpA	Outer membrane protein A	OM	=	$>5.00 \times 10^{-2}$
P39180	Ag43	Antigen 43	OM	=	$>5.00 \times 10^{-2}$
P0A927	Tsx	Nucleoside-specific channel-forming protein Tsx	OM	=	$>5.00 \times 10^{-2}$
P0A940	BamA	Outer membrane protein assembly factor A (Omp85, YaeT)	OM	=	$>5.00 \times 10^{-2}$
Outer membrane lipoproteins (OMLs)					
P69776	Lpp	Major outer membrane lipoprotein Lpp	OM	=	$>5.00 \times 10^{-2}$
P0A905	Pcp	Pal cross-reactive protein, Outer membrane lipoprotein SlyB	OM	=	$>5.00 \times 10^{-2}$
P0ADA5	YajG	Uncharacterised lipoprotein YajG	OM	=	$>5.00 \times 10^{-2}$
P46130	YbhC	Putative acyl-CoA thioester hydrolase YbhC	OM	=	$>5.00 \times 10^{-2}$
P41052	MltB	Membrane-bound lytic murein transglycosylase B	OM	=	$>5.00 \times 10^{-2}$
P64596	YraP	Uncharacterised protein YraP	OM	=	$>5.00 \times 10^{-2}$
P77774	BamB	Outer membrane protein assembly factor B (YfgL)	OM	=	$>5.00 \times 10^{-2}$
P0A903	BamC	Outer membrane protein assembly factor C (NlpB)	OM	=	$>5.00 \times 10^{-2}$
P0A937	BamE	Outer membrane protein assembly factor E (SmpA)	OM	=	$>5.00 \times 10^{-2}$
P0A912	Pal	Peptidoglycan-associated lipoprotein	OM	-2.6	1.65×10^{-2}
P0AC02	BamD	Outer membrane protein assembly factor D (YfiO)	OM	-2.7	1.79×10^{-2}
P0ADK0	YiaF	Uncharacterised protein YiaF	OM	-3.8	1.82×10^{-2}
P45464	LpoA	Penicillin-binding protein activator (YraM)	OM	-5.4	2.96×10^{-2}
Periplasmic proteins					
P08506	DacC	D-alanyl-D-alanine carboxypeptidase	PE	=	$>5.00 \times 10^{-2}$

P0AEG4	DsbA	Thiol:disulfide interchange protein	PE	=	>5.00 x 10^{-2}
P05458	Ptr	Protease 3	PE	=	>5.00 x 10^{-2}
P0C0V0	DegP	Periplasmic serine endoprotease	PE	=	>5.00 x 10^{-2}
P0AE22	AphA	Class B acid phosphatase	PE	=	>5.00 x 10^{-2}
P45523	FkpA	FKBP-type peptidyl-prolyl cis-trans isomerase	PE	=	>5.00 x 10^{-2}
P16700	CysP	Thiosulfate-binding protein	PE	=	>5.00 x 10^{-2}
P0AEX9	MalE	Maltose-binding periplasmic protein	PE	=	>5.00 x 10^{-2}
P45955	YbgF	Cell division coordinator CpoB	PE	=	>5.00 x 10^{-2}
P31697	FimC	Chaperone protein FimC	PE	=	>5.00 x 10^{-2}
Q46863	YgiS	Probable deoxycholate-binding periplasmic protein	PE	=	>5.00 x 10^{-2}
P0AEG6	DsbC	Thiol:disulfide interchange protein	PE	=	>5.00 x 10^{-2}
P0AFH8	OsmY	Osmotically-inducible protein Y	PE	=	>5.00 x 10^{-2}
P0A908	MipA	MltA-interacting protein	PE	=	>5.00 x 10^{-2}
P75797	YliB	Glutathione-binding protein GsiB	PE	=	>5.00 x 10^{-2}
P0AG82	PstS	Phosphate-binding protein PstS (PBP)	PE	=	>5.00 x 10^{-2}
P77202	DsbG	Thiol:disulfide interchange protein DsbG	PE	-1.7	2.08 x 10^{-2}
P23843	OppA	Periplasmic oligopeptide-binding protein	PE	-2.0	6.05 x 10^{-3}
P0AEU7	Skp	Chaperone protein Skp	PE	-2.1	2.59 x 10^{-2}
P39099	DegQ	Periplasmic pH-dependent serine endoprotease	PE	-2.2	1.14 x 10^{-2}
P0ADV7	YrbC	Probable phospholipid-binding protein MlaC	PE	-2.3	1.55 x 10^{-2}
P19926	Agp	Glucose-1-phosphatase (G1Pase)	PE	-2.6	4.05 x 10^{-2}
P23847	DppA	Periplasmic dipeptide transport protein	PE	-2.6	3.01 x 10^{-2}
P76177	YdgH	Protein YdgH	PE	-2.6	1.32 x 10^{-3}
P33363	BglX	Periplasmic beta-glucosidase	PE	-2.6	1.30 x 10^{-2}
P0AB38	YcfM	Penicillin-binding protein activator LpoB	PE	-2.7	7.94 x 10^{-3}
P32176	FdoG	Formate dehydrogenase-O major subunit	PE	-2.7	1.51 x 10^{-2}
P37329	ModA	Molybdate-binding periplasmic protein	PE	-2.8	6.40 x 10^{-4}
P0AEE5	MglB	D-galactose-binding periplasmic protein	PE	-2.8	2.77 x 10^{-3}
P77348	MppA	Periplasmic murein peptide-binding protein	PE	-2.9	5.68 x 10^{-3}
P0AFK9	PotD	Spermidine/putrescine-binding periplasmic protein	PE	-3.1	4.02 x 10^{-3}
P09394	GlpQ	Glycerophosphodiester phosphodiesterase	PE	-3.1	3.45 x 10^{-2}
P23865	Prc	Tail-specific protease	PE	-3.2	1.41 x 10^{-2}

P03841	MalM	Maltose operon periplasmic protein	PE	-3.3	6.02×10^{-3}
P33136	MdoG	Glucans biosynthesis protein G	PE	-3.4	2.30×10^{-4}
P0AEM9	FliY	L-cystine-binding protein FliY	PE	-3.4	2.30×10^{-2}
P40120	MdoD	Glucans biosynthesis protein D	PE	-3.7	1.97×10^{-2}
P0A855	TolB	Protein TolB	PE	-3.8	1.80×10^{-5}
P76116	YncE	Uncharacterised protein YncE	PE	-3.8	6.50×10^{-4}
P07024	UshA	Protein UshA	PE	-3.8	1.41×10^{-3}
P02925	RbsB	Ribose import binding protein	PE	-4.2	8.70×10^{-3}
P0AEQ3	GlnH	Glutamine-binding periplasmic protein	PE	-4.4	3.57×10^{-3}
P77318	YdeN	Uncharacterised sulfatase YdeN	PE	-5.0	1.81×10^{-2}
P0ABZ6	SurA	Chaperone SurA	PE	-5.3	5.30×10^{-4}
P39325	YtfQ	ABC transporter periplasmic-binding protein YtfQ	PE	-5.6	2.54×10^{-2}
P0AFX9	RseB	Sigma-E factor regulatory protein	PE	-7.2	2.46×10^{-2}
P23827	Eco	Ecotin	PE	-20.2	1.32×10^{-3}
P0AAX8	YbiS	Periplasmic L,D-transpeptidase	PE	-	1.20×10^{-6}
P37902	GltI	Glutamate/aspartate import solute-binding protein	PE	-	2.00×10^{-6}
P0AEU0	HisJ	Histidine-binding periplasmic protein (HBP)	PE	-	1.90×10^{-5}
Inner membrane proteins (IMPs)					
P0ABB0	AtpA	ATP synthase subunit α	IM	+2.9	1.22×10^{-2}
P0ABB4	AtpD	ATP synthase subunit β	IM	+2.4	4.60×10^{-4}
P0ABA6	AtpG	ATP synthase gamma chain	IM	+2.2	4.51×10^{-2}
P46889	FtsK	DNA translocase FtsK	IM	=	$>5.00 \times 10^{-2}$
P06959	AceF	Dihydrolipoyllysine-residue acetyltransferase	IM	=	$>5.00 \times 10^{-2}$
P0AC41	SdhA	Succinate dehydrogenase flavoprotein subunit	IM	=	$>5.00 \times 10^{-2}$
P0AFG8	AceE	Pyruvate dehydrogenase E1 component	IM	=	$>5.00 \times 10^{-2}$
P69797	ManX	PTS system mannose-specific EIIAB component	IM	=	$>5.00 \times 10^{-2}$
P07014	SdhB	Succinate dehydrogenase iron-sulfur subunit	IM	=	$>5.00 \times 10^{-2}$
P33602	NuoG	NADH-quinone oxidoreductase subunit G	IM	=	$>5.00 \times 10^{-2}$
P0AAJ5	FdoH	Formate dehydrogenase-O iron-sulfur subunit	IM	=	$>5.00 \times 10^{-2}$
P0ABJ9	CydA	Cytochrome bd-I ubiquinol oxidase subunit 1	IM	=	$>5.00 \times 10^{-2}$
P0ABC3	HflC	Modulator of FtsH protease HflC	IM	=	$>5.00 \times 10^{-2}$
P25714	YidC	Membrane protein insertase YidC	IM	=	$>5.00 \times 10^{-2}$

P76372	Cld	Chain length determinant protein	IM	=	>5.00 x 10^{-2}
P0AEL0	FdoI	Formate dehydrogenase. cytochrome b556(fdo) subunit	IM	=	>5.00 x 10^{-2}
P45757	T2SC	Type II secretion system protein C (T2SS protein C)	IM	=	>5.00 x 10^{-2}
P31224	AcrB	Multidrug efflux pump subunit	IM	=	>5.00 x 10^{-2}
P08395	SppA	Protease 4	IM	=	>5.00 x 10^{-2}
P0ADV9	YrbK	Lipopolysaccharide export system protein	IM	=	>5.00 x 10^{-2}
P0ADT8	HtrG	Uncharacterised protein YgiM	IM	=	>5.00 x 10^{-2}
P37626	YhiI	Uncharacterised protein YhiI	IM	=	>5.00 x 10^{-2}
P0ABI8	CyoB	Cytochrome bo(3) ubiquinol oxidase subunit 1	IM	=	>5.00 x 10^{-2}
P77338	KefA	Mechanosensitive channel MscK	IM	=	>5.00 x 10^{-2}
P0ADZ7	YajC	UPF0092 membrane protein YajC	IM	=	>5.00 x 10^{-2}
P15078	CstA	Carbon starvation protein A	IM	=	>5.00 x 10^{-2}
P0ABA0	AtpF	ATP synthase subunit b	IM	=	>5.00 x 10^{-2}
P07017	Tar	Methyl-accepting chemotaxis protein II	IM	=	>5.00 x 10^{-2}
P69831	GatC	PTS system galactitol-specific EIIC component	IM	=	>5.00 x 10^{-2}
P36672	TreB	PTS system trehalose-specific EIIBC component	IM	=	>5.00 x 10^{-2}
P33650	FeoB	Fe(2+) transporter	IM	=	>5.00 x 10^{-2}
P05704	Trg	Methyl-accepting chemotaxis protein III	IM	=	>5.00 x 10^{-2}
P0AAD6	SdaC	Serine transporter	IM	=	>5.00 x 10^{-2}
P69801	ManY	PTS system mannose-specific EIIC component	IM	=	>5.00 x 10^{-2}
P02942	Tsr	Methyl-accepting chemotaxis protein I	IM	=	>5.00 x 10^{-2}
P0ABN5	DcuA	Anaerobic C4-dicarboxylate transporter	IM	=	>5.00 x 10^{-2}
P0ABC7	HflK	Modulator of FtsH protease HflK	IM	-1.6	3.19 x 10^{-2}
P64604	MlaD	Probable phospholipid ABC transporter-binding protein	IM	-1.7	2.47 x 10^{-2}
P0ABX8	FliL	Flagellar protein FliL	IM	-1.7	4.63 x 10^{-2}
P39396	YjiY	Inner membrane protein YjiY	IM	-2.0	2.65 x 10^{-3}
P77804	YdgA	Protein YdgA	IM	-2.0	1.95 x 10^{-2}
P0ADY1	PpiD	Peptidyl-prolyl cis-trans isomerase D	IM	-2.2	1.20 x 10^{-4}
P0ADS6	YggE	Uncharacterised protein YggE	IM	-2.3	3.06 x 10^{-3}
P0AAI3	FtsH	ATP-dependent zinc metalloprotease FtsH	IM	-2.3	4.54 x 10^{-2}
P11557	DamX	Cell division protein DamX	IM	-2.5	2.45 x 10^{-2}
P0AEB2	DacA	D-alanyl-D-alanine carboxypeptidase DacA	IM	-2.6	2.49 x 10^{-2}

P09127	HemX	Putative uroporphyrinogen-III C-methyltransferase	IM	-3.2	6.36×10^{-3}
P39401	MdoB	Phosphoglycerol transferase I	IM	-6.2	2.35×10^{-2}
P07018	Tap	Methyl-accepting chemotaxis protein IV	IM	-	1.60×10^{-8}
Inner membrane lipoproteins (IMLs)					
P76578	YfhM	Uncharacterised lipoprotein YfhM	IM	=	$>5.00 \times 10^{-2}$
P0ADA3	NlpD	Murein hydrolase activator	IM	=	$>5.00 \times 10^{-2}$
P0AE06	AcrA	Multidrug efflux pump subunit	IM	-2.5	7.92×10^{-3}
Cell-surface moonlighting proteins					
P60723	RplD	50S ribosomal protein L4	CY	+1.7	2.93×10^{-2}
P0A7L0	RplA	50S ribosomal protein L1	CY	=	$>5.00 \times 10^{-2}$
P0A7V8	RpsD	30S ribosomal protein S4	CY	=	$>5.00 \times 10^{-2}$
P0A7K2	RplL	50S ribosomal protein L7/L12 (L8)	CY	=	$>5.00 \times 10^{-2}$
P0A7J3	RplJ	50S ribosomal protein L10	CY	=	$>5.00 \times 10^{-2}$
P0A9B2	GapA	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	CY	=	$>5.00 \times 10^{-2}$
P0A6P9	Eno	Enolase	CY	=	$>5.00 \times 10^{-2}$
P0A799	Pgk	Phosphoglycerate kinase	CY	=	$>5.00 \times 10^{-2}$
P0AD61	PykF	Pyruvate kinase	CY	=	$>5.00 \times 10^{-2}$
P0A858	TpiA	Triosephosphate isomerase	CY	=	$>5.00 \times 10^{-2}$
P08200	Icd	Isocitrate dehydrogenase (IDH)	CY	=	$>5.00 \times 10^{-2}$
P0ABH7	GltA	Citrate synthase	CY	=	$>5.00 \times 10^{-2}$
P0A9Q7	AdhE	Aldehyde-alcohol dehydrogenase	CY	=	$>5.00 \times 10^{-2}$
P00350	Gnd	6-phosphogluconate dehydrogenase	CY	=	$>5.00 \times 10^{-2}$
P0ABK5	CysK	Cysteine synthase A	CY	=	$>5.00 \times 10^{-2}$
P0A9C5	GlnA	Glutamine synthetase (GS)	CY	=	$>5.00 \times 10^{-2}$
P00448	SodA	Superoxide dismutase	CY	=	$>5.00 \times 10^{-2}$
P0A825	GlyA	Serine hydroxymethyltransferase	CY	=	$>5.00 \times 10^{-2}$
P0A6Y8	DnaK	Chaperone protein DnaK (HSP70)	CY	=	$>5.00 \times 10^{-2}$
P0A6F5	GroL	60 kDa chaperonin	CY	=	$>5.00 \times 10^{-2}$
P0CE47	TufA	Elongation factor Tu (EF-Tu)	CY	=	$>5.00 \times 10^{-2}$
P0A6M8	FusA	Elongation factor G (EF-G)	CY	=	$>5.00 \times 10^{-2}$
P0A8V2	RpoB	DNA-directed RNA polymerase subunit β	CY	=	$>5.00 \times 10^{-2}$

[a] Predicted subcellular localisation based on the database of Díaz-Mejía *et al*. [29], OM: outer membrane (GO:0036406), PE: periplasm (GO: 0042597), IM: inner membrane (GO: 0005886), CY: cytoplasm (GO: 0005737).

[b] Fold change corresponds to the differential protein abundance following biotinylation with high-molecular weight sulfo-NHS-PEG4-bismannose-SS-biotin versus low-molecular weight sulfo-NHS-SS-biotin. Fold change is presented with positive value (+) when the proteins was significantly higher abundant with sulfo-NHS-PEG4-bismannose-SS-biotin and with a negative value (-) when the protein was lower abundant with sulfo-NHS-PEG4-bismannose-SS-biotin compared to biotinylation with sulfo-NHS-SS-biotin used as the reference condition. Differences were considered significant for *p*-values below 0.05 (supplementary material Table 1S).

ACKNOWLEDGEMENTS

This work was supported in part by INRA (Institut National de la Recherche Agronomique) and by EU FP7 Marie Curie Actions ITN (Initial Training Network) EID (European Industrial Doctorates) DIS*Co* (A multidisciplinary Doctoral Industrial School on novel preventive strategies against *E. Coli* infections; http://www.discoproject.eu/; n°FP7-PEOPLE-607611) project. Ricardo MONTEIRO is a Marie Curie PhD Research Fellow granted by ITN EID DIS*Co*. The authors thank Hicham BESSAIAH and Florelle DEBOUDARD for technical assistance during their MSc traineeships.

SUPPLEMENTARY MATERIAL

Table 1S: Proteins identified through biotin labelling using sulfo-NHS-SS-biotin and sulfo-NHS-PEG4-bismannose-SS-biotin.

REFERENCES

1. Thanassi DG, Bliska JB, Christie PJ. Surface organelles assembled by secretion systems of Gram-negative bacteria: diversity in structure and function. *FEMS microbiology reviews* 2012; **36**: 1046-82.
2. Lee VT, Schneewind O. Protein secretion and the pathogenesis of bacterial infections. *Genes Dev* 2001; **15**: 1725-52.
3. Stubbendieck RM, Straight PD. Multifaceted Interfaces of Bacterial Competition. *J Bacteriol* 2016; **198**: 2145-55.
4. Rediers H, Rainey PB, Vanderleyden J, De Mot R. Unraveling the secret lives of bacteria: use of in vivo expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. *Microbiol Mol Biol Rev* 2005; **69**: 217-61.
5. Chagnot C, Zorgani MA, Astruc T, Desvaux M. Proteinaceous determinants of surface colonisation in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front Microbiol* 2013; **4**: 303.
6. Desvaux M, Hebraud M, Talon R, Henderson IR. Secretion and subcellular localisations of bacterial proteins: a semantic awareness issue. *Trends Microbiol* 2009; **17**: 139-45.
7. Delepelaire P, Wandersman C. Protein Export and Secretion in Gram-Negative Bacteria. *Microbial Transport Systems*: Wiley-VCH Verlag GmbH & Co. KGaA 2003:165-208.
8. Tan S, Tan HT, Chung MC. Membrane proteins and membrane proteomics. *Proteomics* 2008; **8**: 3924-32.
9. Cordwell SJ. Technologies for bacterial surface proteomics. *Curr Opin Microbiol* 2006; **9**: 320-9.
10. Solis N, Cordwell SJ. Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics* 2011; **11**: 3169-89.
11. Santoni V, Molloy M, Rabilloud T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* 2000; **21**: 1054-70.
12. Thein M, Sauer G, Paramasivam N et al. Efficient subfractionation of gram-negative bacteria for proteomics studies. *J Proteome Res* 2010; **9**: 6135-47.
13. Walters MS, Mobley HL. Identification of uropathogenic *Escherichia coli* surface proteins by shotgun proteomics. *J Microbiol Methods* 2009; **78**: 131-5.
14. Cullen PA, Xu X, Matsunaga J et al. Surfaceome of *Leptospira* spp. *Infect Immun* 2005; **73**: 4853-63.
15. Scheurer SB, Roesli C, Neri D, Elia G. A comparison of different biotinylation reagents, tryptic digestion procedures, and mass spectrometric techniques for 2-D peptide mapping of membrane proteins. *Proteomics* 2005; **5**: 3035-9.
16. Bradburne JA, Godfrey P, Choi JH, Mathis JN. In vivo labeling of *Escherichia coli* cell envelope proteins with N-hydroxysuccinimide esters of biotin. *Applied and environmental microbiology* 1993; **59**: 663-8.
17. Sabarth N, Lamer S, Zimny-Arndt U et al. Identification of surface proteins of *Helicobacter pylori* by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. *J Biol Chem* 2002; **277**: 27896-902.
18. Pornwiroon W, Bourchookarn A, Paddock CD, Macaluso KR. Proteomic analysis of *Rickettsia parkeri* strain portsmouth. *Infect Immun* 2009; **77**: 5262-71.

19. Smither SJ, Hill J, van Baar BL et al. Identification of outer membrane proteins of *Yersinia pestis* through biotinylation. *J Microbiol Methods* 2007; **68**: 26-31.
20. Madler S, Bich C, Touboul D, Zenobi R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *J Mass Spectrom* 2009; **44**: 694-706.
21. Shimkus M, Levy J, Herman T. A chemically cleavable biotinylated nucleotide: usefulness in the recovery of protein-DNA complexes from avidin affinity columns. *Proc Natl Acad Sci U S A* 1985; **82**: 2593-7.
22. Elia G. Biotinylation reagents for the study of cell surface proteins. *Proteomics* 2008; **8**: 4012-24.
23. Ge Y, Rikihisa Y. Surface-exposed proteins of *Ehrlichia chaffeensis*. *Infect Immun* 2007; **75**: 3833-41.
24. Ge Y, Rikihisa Y. Identification of novel surface proteins of *Anaplasma phagocytophilum* by affinity purification and proteomics. *J Bacteriol* 2007; **189**: 7819-28.
25. Gibson K, Kumagai Y, Rikihisa Y. Proteomic analysis of *Neorickettsia sennetsu* surface-exposed proteins and porin activity of the major surface protein P51. *J Bacteriol* 2010; **192**: 5898-905.
26. Zhang H, Brown RN, Qian WJ et al. Quantitative analysis of cell surface membrane proteins using membrane-impermeable chemical probe coupled with 18O labeling. *J Proteome Res* 2010; **9**: 2160-9.
27. Gong W, Xiong X, Qi Y et al. Identification of novel surface-exposed proteins of *Rickettsia rickettsii* by affinity purification and proteomics. *PloS one* 2014; **9**: e100253.
28. Sears KT, Ceraul SM, Gillespie JJ et al. Surface proteome analysis and characterisation of surface cell antigen (Sca) or autotransporter family of *Rickettsia typhi*. *PLoS Pathog* 2012; **8**: e1002856.
29. Qi Y, Xiong X, Wang X et al. Proteome analysis and serological characterisation of surface-exposed proteins of *Rickettsia heilongjiangensis*. *PloS one* 2013; **8**: e70440.
30. Wilson MM, Anderson DE, Bernstein HD. Analysis of the outer membrane proteome and secretome of *Bacteroides fragilis* reveals a multiplicity of secretion mechanisms. *PloS one* 2015; **10**: e0117732.
31. Benz R, Bauer K. Permeation of hydrophilic molecules through the outer membrane of gram-negative bacteria. Review on bacterial porins. *Eur J Biochem* 1988; **176**: 1-19.
32. Diaz-Mejia JJ, Babu M, Emili A. Computational and experimental approaches to chart the *Escherichia coli* cell-envelope-associated proteome and interactome. *FEMS microbiology reviews* 2009; **33**: 66-97.
33. Bhatia VN, Perlman DH, Costello CE, McComb ME. Software tool for researching annotations of proteins: open-source protein annotation software with data visualisation. *Anal Chem* 2009; **81**: 9819-23.
34. Hempel K, Pane-Farre J, Otto A et al. Quantitative cell surface proteome profiling for SigB-dependent protein expression in the human pathogen *Staphylococcus aureus* via biotinylation approach. *J Proteome Res* 2010; **9**: 1579-90.
35. Hempel K, Herbst FA, Moche M et al. Quantitative proteomic view on secreted, cell surface-associated, and cytoplasmic proteins of the methicillin-resistant human pathogen *Staphylococcus aureus* under iron-limited conditions. *J Proteome Res* 2011; **10**: 1657-66.
36. Desvaux M, Hebraud M, Henderson IR, Pallen MJ. Type III secretion: what's in a name? *Trends Microbiol* 2006; **14**: 157-60.

37. Macnab RM. Type III flagellar protein export and flagellar assembly. *Biochim Biophys Acta* 2004; **1694**: 207-17.
38. Koronakis V, Eswaran J, Hughes C. Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annu Rev Biochem* 2004; **73**: 467-89.
39. Confer AW, Ayalew S. The OmpA family of proteins: roles in bacterial pathogenesis and immunity. *Vet Microbiol* 2013; **163**: 207-22.
40. van der Woude MW, Henderson IR. Regulation and function of Ag43 (flu). *Annu Rev Microbiol* 2008; **62**: 153-69.
41. Selkrig J, Leyton DL, Webb CT, Lithgow T. Assembly of beta-barrel proteins into bacterial outer membranes. *Biochim Biophys Acta* 2014; **1843**: 1542-50.
42. Parsons LM, Lin F, Orban J. Peptidoglycan recognition by Pal, an outer membrane lipoprotein. *Biochemistry* 2006; **45**: 2122-8.
43. Jean NL, Bougault CM, Lodge A et al. Elongated structure of the outer-membrane activator of peptidoglycan synthesis LpoA: implications for PBP1A stimulation. *Structure* 2014; **22**: 1047-54.
44. Chang TW, Lin YM, Wang CF, Liao YD. Outer membrane lipoprotein Lpp is Gram-negative bacterial cell surface receptor for cationic antimicrobial peptides. *J Biol Chem* 2012; **287**: 418-28.
45. Lo M, Boyce JD, Wilkie IW, Adler B. Characterisation of two lipoproteins in *Pasteurella multocida*. *Microbes Infect* 2004; **6**: 58-67.
46. Kim KH, Aulakh S, Paetzel M. Crystal structure of beta-barrel assembly machinery BamCD protein complex. *J Biol Chem* 2011; **286**: 39116-21.
47. Henderson B, Martin A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun* 2011; **79**: 3476-91.
48. Egea L, Aguilera L, Gimenez R et al. Role of secreted glyceraldehyde-3-phosphate dehydrogenase in the infection mechanism of enterohemorrhagic and enteropathogenic *Escherichia coli:* interaction of the extracellular enzyme with human plasminogen and fibrinogen. *Int J Biochem Cell Biol* 2007; **39**: 1190-203.
49. Pancholi V. Multifunctional alpha-enolase: its role in diseases. *Cell Mol Life Sci* 2001; **58**: 902-20.
50. Boel G, Pichereau V, Mijakovic I et al. Is 2-phosphoglycerate-dependent automodification of bacterial enolases implicated in their export? *J Mol Biol* 2004; **337**: 485-96.
51. Amimanan P, Tavichakorntrakool R, Fong-Ngern K et al. Elongation factor Tu on *Escherichia coli* isolated from urine of kidney stone patients promotes calcium oxalate crystal growth and aggregation. *Sci Rep* 2017; **7**: 2953.
52. Gatlin CL, Pieper R, Huang ST et al. Proteomic profiling of cell envelope-associated proteins from *Staphylococcus aureus*. *Proteomics* 2006; **6**: 1530-49.
53. Romero-Saavedra F, Laverde D, Wobser D et al. Identification of peptidoglycan-associated proteins as vaccine candidates for enterococcal infections. *PloS one* 2014; **9**: e111880.
54. Becher D, Hempel K, Sievers S et al. A proteomic view of an important human pathogen--towards the quantification of the entire *Staphylococcus aureus* proteome. *PloS one* 2009; **4**: e8176.
55. Nikaido H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* 2003; **67**: 593-656.

56. Nikaido H, Vaara M. Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* 1985; **49**: 1-32.
57. Stokes JM, French S, Ovchinnikova OG et al. Cold Stress Makes *Escherichia coli* Susceptible to Glycopeptide Antibiotics by Altering Outer Membrane Integrity. *Cell Chem Biol* 2016; **23**: 267-77.
58. Elia G. Protein biotinylation. *Curr Protoc Protein Sci* 2010; **Chapter 3**: Unit 3 6.
59. Pothula KR, Solano CJ, Kleinekathofer U. Simulations of outer membrane channels and their permeability. *Biochim Biophys Acta* 2016; **1858**: 1760-71.
60. Kainulainen V, Korhonen TK. Dancing to another tune-adhesive moonlighting proteins in bacteria. *Biology (Basel)* 2014; **3**: 178-204.
61. Sanchez B, Schmitter JM, Urdaci MC. Identification of novel proteins secreted by *Lactobacillus rhamnosus* GG grown in de Mann-Rogosa-Sharpe broth. *Lett Appl Microbiol* 2009; **48**: 618-22.
62. Boone TJ, Burnham CA, Tyrrell GJ. Binding of group B streptococcal phosphoglycerate kinase to plasminogen and actin. *Microb Pathog* 2011; **51**: 255-61.
63. Boone TJ, Tyrrell GJ. Identification of the actin and plasminogen binding regions of group B streptococcal phosphoglycerate kinase. *J Biol Chem* 2012; **287**: 29035-44.
64. Ramiah K, van Reenen CA, Dicks LM. Surface-bound proteins of *Lactobacillus plantarum* 423 that contribute to adhesion of Caco-2 cells and their role in competitive exclusion and displacement of *Clostridium sporogenes* and *Enterococcus faecalis*. *Res Microbiol* 2008; **159**: 470-5.
65. Monteiro R, Ageorges V, Rojas-Lopez M et al. A secretome view of colonisation factors in Shiga toxin-encoding *Escherichia coli* (STEC): from enterohaemorrhagic *E. coli* (EHEC) to related enteropathotypes. *FEMS Microbiol Lett* 2016; **363**.
66. Renier S, Micheau P, Talon R et al. Subcellular localisation of extracytoplasmic proteins in monoderm bacteria: rational secretomics-based strategy for genomic and proteomic analyses. *PloS one* 2012; **7**: e42982.
67. Serruto D, Rappuoli R. Post-genomic vaccine development. *FEBS Lett* 2006; **580**: 2985-92.
68. Molloy MP, Herbert BR, Slade MB et al. Proteomic analysis of the *Escherichia coli* outer membrane. *Eur J Biochem* 2000; **267**: 2871-81.
69. Fountoulakis M, Gasser R. Proteomic analysis of the cell envelope fraction of *Escherichia coli*. *Amino Acids* 2003; **24**: 19-41.
70. Lopez-Campistrous A, Semchuk P, Burke L et al. Localisation, annotation, and comparison of the *Escherichia coli* K-12 proteome under two states of growth. *Mol Cell Proteomics* 2005; **4**: 1205-9.
71. Piatigorsky J, Wistow GJ. Enzyme/crystallins: gene sharing as an evolutionary strategy. *Cell* 1989; **57**: 197-9.
72. Aguilera L, Gimenez R, Badia J et al. NAD+-dependent post-translational modification of *Escherichia coli* glyceraldehyde-3-phosphate dehydrogenase. *Int Microbiol* 2009; **12**: 187-92.
73. Knaust A, Weber MV, Hammerschmidt S et al. Cytosolic proteins contribute to surface plasminogen recruitment of *Neisseria meningitidis*. *J Bacteriol* 2007; **189**: 3246-55.
74. Dallo SF, Kannan TR, Blaylock MW, Baseman JB. Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in Mycoplasma pneumoniae. *Mol Microbiol* 2002; **46**: 1041-51.
75. Singh D, Chang SJ, Lin PH et al. Regulation of ribonuclease E activity by the L4 ribosomal protein of *Escherichia coli*. *Proc Natl Acad Sci U S A* 2009; **106**: 864-9.

76. Mani M, Chen C, Amblee V et al. MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res* 2015; **43**: D277-82.
77. Werts C, O'Callaghan D, Hofnung M, Charbit A. Immunological relatedness of the LamB proteins among members of Enterobacteriaceae. *J Gen Microbiol* 1993; **139**: 881-7.
78. Gilson E, Rousset JP, Charbit A et al. malM, a new gene of the maltose regulon in *Escherichia coli* K12. I. malM is the last gene of the malK-lamB operon and encodes a periplasmic protein. *J Mol Biol* 1986; **191**: 303-11.
79. Rousset JP, Gilson E, Hofnung M. malM, a new gene of the maltose regulon in *Escherichia coli K12*. II. Mutations affecting the signal peptide of the MalM protein. *J Mol Biol* 1986; **191**: 313-20.
80. Macnab RM. How bacteria assemble flagella. *Annu Rev Microbiol* 2003; **57**: 77-100.
81. Schoenhals GJ, Macnab RM. FliL is a membrane-associated component of the flagellar basal body of *Salmonella*. *Microbiology* 1999; **145 (Pt 7)**: 1769-75.
82. Bonifield HR, Yamaguchi S, Hughes KT. The flagellar hook protein, FlgE, of *Salmonella enterica* serovar typhimurium is posttranscriptionally regulated in response to the stage of flagellar assembly. *J Bacteriol* 2000; **182**: 4044-50.
83. He Y, Xu T, Fossheim LE, Zhang XH. FliC, a flagellin protein, is essential for the growth and virulence of fish pathogen *Edwardsiella tarda*. *PloS one* 2012; **7**: e45070.
84. Paradis-Bleau C, Markovski M, Uehara T et al. Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. *Cell* 2010; **143**: 1110-20.
85. Godlewska R, Wisniewska K, Pietras Z, Jagusztyn-Krynicka EK. Peptidoglycan-associated lipoprotein (Pal) of Gram-negative bacteria: function, structure, role in pathogenesis and potential application in immunoprophylaxis. *FEMS Microbiol Lett* 2009; **298**: 1-11.

CHAPTER IV: THE EXTRACYTOPLASMIC PROTEOME LANDSCAPE OF *ESCHERICHIA COLI* O157:H7: FROM CELL-SURFACE, EXTRACELLULAR TO OUTER MEMBRANE VESICLES SUBPROTEOMES OF A PATHOGENIC LIPOPOLYSACCHARIDIC DIDERM BACTERIUM

The Extracytoplasmic Proteome Landscape of *Escherichia coli* O157:H7: from Cell-Surface, Extracellular to Outer Membrane Vesicles Subproteomes of a Pathogenic Lipopolysaccharidic Diderm Bacterium

Ricardo MONTEIRO [1,2], Ingrid CHAFSEY [1], Sabine LEROY [1], Christophe CHAMBON [3], Michel HÉBRAUD [1,3], Valérie LIVRELLI [4], Mariagrazia PIZZA [2], Alfredo PEZZICOLI [2], Mickaël DESVAUX [1,*]

[1] Université Clermont Auvergne, INRA, UMR454 MEDiS, F-63000 Clermont-Ferrand, France.

[2] GSK Vaccines, Via Fiorentina 1, 53100, Siena, Italy.

[3] INRA, Plate-Forme d'Exploration du Métabolisme, F-63122 Saint-Genès Champanelle, France.

[4] Centre de Recherche en Nutrition Humaine Auvergne, UMR UCA INSERM U1071, USC-INRA 2018, Clermont Université - Université d'Auvergne, Faculté de Pharmacie, CHU Clermont-Ferrand, Service Bactériologie Mycologie Parasitologie, Clermont-Ferrand, France.

*Corresponding author: Dr Hab Mickaël DESVAUX-LENÔTRE, INRA (Institut National de la Recherche Agronomique) Auvergne-Rhône-Alpes, Centre de Recherche de Clermont-Ferrand, UMR454 MEDiS (Microbiologie, Environnement Digestif, Santé), Site de Theix, F-63122 Saint-Genès Champanelle, France. Tel.: +33 (0)4 73 62 47 23, Fax: +33 (0)4 73 62 45 81, E-mail: mickael.desvaux@inra.fr.

ABSTRACT

Pathogenic *E. coli* can be broadly classified as either extraintestinal *E. coli* (ExPEC) or diarrheagenic intestinal E. coli (InPEC). Among the six well-described enteropathotypes, enterohaemorrhagic *E. coli* (EHEC) are the most virulent anthropozoonotic agents. The ability of *E. coli* to colonise different environments and to cause a wide range of disease is partly due to its ability to secrete different effectors at the bacterial cell surface or in the extracellular milieu that varies from one strain to another due to the high plasticity of the *E. coli* genome. This work aims to identify the differences on secreted proteins by EHEC O157:H7 grown in different culture conditions, following proteomic analysis of the cell surface, extracellular and outer membrane vesicles proteomes. EHEC O157:H7 strain EDL933 was cultured in different nutrient media, namely DMEM, BHI and M9. Proteome induced changes by culture conditions were evaluated using label-free quantitation protocol adapted from biotinylation method. Quantitative analysis of the three subproteomes revealed unique proteins expression depending on growth conditions and related to different secretion pathways from EHEC O157:H7. Our results showed more abundant proteins secreted by T3aSS (injectisome) when EHEC were grown in DMEM contrasting with BHI where EHEC were secreting more proteins through T3bSS (flagellum). Moreover growing this pathogen in minimal chemical conditions such M9 culture medium, revealed the secretion full potential of bacteria increasin the abundance of proteins secreted by T1SS. This study concluded that the culture condition in which EHEC are pre culture subsequent infection experiments could be a major factor. The identification of those proteins will constitute important molecular targets for the development of preventive strategies against *E. coli* contamination and infection along the food chain. Also the understanding of culture influence in the secretion profile of entero pathogens will allow us to approach the intestinal reality in terms of infection.

INTRODUCTION

Escherichia coli are an inhabitant of warm-blooded animals intestines and typically coexists with the host as a commensal with mutual benefit for both organisms [1]. However there are highly adapted *E. coli* clones that have acquired specific virulence attributes, which allow them to cause a broad spectrum of disease. Based on clinical syndromes, the pathogenic strains can be broadly discriminated into intestinal pathogenic *E. coli* (InPEC) resulting in diarrhea, and extraintestinal pathogenic *E. coli* (ExPEC) responsible for urinary tract infections, sepsis and/or meningitis [2]. InPEC represents a leading cause for traveller's diarrhea but also pediatric disease in developing countries, and an emerging source of diarrhea in industrialised countries [3]. Enterohemorragic *Escherichia coli* (EHEC) O157:H7, subset of Shiga-toxin producing *E.* coli (STEC), represents the major etiologic agent of diseases in humans, with wide spectrum ranging from aqueous or bloody diarrhoea to severe cases such haemorrhagic colitis and haemolytic uremic syndrome (HUS), the leading cause of chronic renal failure in children in several developing countries [4, 5]. Shiga-toxin is the main virulence factor, but the ability of EHEC to colonise the intestinal epithelial surface and cause histopathological alterations at so-called attaching and effacing lesions (A/E lesions) is crucial for the success of EHEC infection [6]. Most of the genes necessary to form A/E lesions are encoded by a 35-kb chromosomal pathogenicity island (PAI) known as the LEE (locus of enterocyte effacement), particularly genes encoding Type III secretion system (T3SS) and its set of associated effector proteins [7, 8]. Actually, intestinal colonisation is a key step in EHEC O157:H7 pathogenesis but it is still not completely understood. Further information on the adherence-related factors expressed *in vivo* is important to decipher EHEC adherence mechanisms [2]. However, the gap of knowledge between infection

in vivo and *in vitro* models remains a giant leap. The classic A/E histopathology has been seen in gnotobiotic piglets [9-12], infant rabbits [13], and cultured epithelial cells [14, 15] infected with *E. coli* O157:H7. Of note, however, A/E lesions are never observed in clinical samples [2]. The ideal conditions to study A/E pathogens should be those existing in the intestine or those mimicking them [16] but performing bacterial studies on those conditions is extremely complex. Therefore, A/E pathogens can be cultured in the laboratory under inducing conditions such DMEM (Dulbecco´s modified Eagle´s Medium): optimal conditions for the expression and secretion of virulence proteins encoded in LEE PAI [16, 17]. This tissue culture medium promotes the efficient adherence to epithelial cells, comparing when they are grown in LB (Lysogeny broth) culture medium [18]. Its known that this feature is due to the presence of sodium bicarbonate, a signalling molecule important to activate virulence determinants required for infection [16]. Also BHI complemented with sodium bicarbonate showed to increase the adhesion capability of EHEC to tissue cultured cells [14]. Expression of the LEE genes and potential adherence factors are under complex regulation and are affected by several factors, such bacterial cell growth rate, temperature, ions, pH, osmolarity [16, 19]. Consequently, such regulation directly affects the coordination and time of the secretion of these important adhesion factors. In bacterial physiology, the use of chemical defined and minimal medium is a classical and efficient way to assay the metabolic capability of a microorganism by forcing it to express its full potential, both in terms of catabolic and anabolic pathways. Protein secretion is an essential anabolic cellular function involved in changes of the extracytoplasmic proteome, i.e. at the bacterial cell-envelope, the cell surface and the extracellular milieu levels. Up to date, nine secretion systems have been uncovered in Gram-negative bacteria and among them, eight may be potentially involved in surface colonisation [20]. While secretion refers to the active transport

from the interior to the exterior of the cell compartment, secreted proteins have three main possible fates in LPS-diderm bacteria: (i) remain associated with the bacterial outer membrane (OM), (ii) be released into the extracellular milieu, (iii) injected into a target cell (either a eukaryotic or bacterial cell) or (iv) through release of outer membrane vesicles [21, 22]. Yet, experimental access to the cell-surface proteome (proteosurfaceome) has been proven difficult. In fact, disclosing this sub proteome in LPS-diderm bacteria by classical proteomic approaches is impaired by numerous factors, such as the hydrophobicity of outer membrane proteins (OMPs) or the low abundance of cell surface proteins [23, 24]. Influence of *in vitro* culture conditions on protein secretion and adherence of EHEC to posterior *in vivo* intestinal infection models has not been clarified.

The present study aims at deciphering the extracytoplasmic protein expression landscape by using a range a growth media, already known and/or most likely to induce changes in the expression of virulence factors in *E. coli* O157:H7, namely DMEM, BHI and M9 media. To get a full picture of EHEC secretome, three subproteomes were here considered and characterised by relative quantification analysis, namely, the proteosurfaceome (proteome at the cell surface), exoproteome (proteome in the extracellular milieu) and proteovesiculome (proteome of the membrane vesicles).

MATERIAL AND METHODS

Bacterial growth conditions

Escherichia coli O157:H7 strain EDL 933 [25] was used for this study. Pre-cultures were grown overnight in DMEM, BHI and M9 liquid media at 37 °C under orbital shaking in triplicates. After 1:100 (v/v) dilution, bacterial culture were grown in the same conditions until late exponential phase (0.8 $OD_{600\ nm}$). Bacterial cells were harvest by centrifugation (4000 g, 5 min, 4 °C) and washed twice with 10 mM PBS pH 8 (4000 g, 5 min, 4 °C). Respective supernatants where kept for further recovery of extracellular proteins and outer membrane vesicules.

Biotinylation of bacterial cell surface proteins and NeutrAvidin affinity purification

Bacterial cells were harvest by centrifugation (4000 g, 5 min, 4 °C) and washed twice with 10 mM PBS pH 8 (4000 g, 5 min, 4 °C). Pelleted cells were weighed and resuspended in 10 mM PBS pH 8. Intact EHEC bacterial cells were incubated with Sulfo-NHS-SS-biotin (Sulfo-succinimidyl biotin-amidoethyl dithio-propionate ; 606.69 Da; Thermo Scientific) at 1 % (m/m), during 1 min with gentle agitation at room temperature. Excess of biotinylation reagent was quenched by three washes with a solution of 10 mM PBS, pH 8 and 500 mM Glycine (4000 g, 5 min, 4°C) and cells were resuspended in lysis buffer (10 mM PBS pH 8, 1 mM PMSF, 1 % v/v TritionX100). As reaction control, cells were incubated with 10 mM PBS, pH 8 instead of biotin reagent and underwent the same procedure. Cell disruption was performed using Fast-prep (MP Biomedicals) with two steps of 20 seconds at 6 m/s and the cell debris from cell lysis discarded by centrifugation (20 000 g, 30 min, 4 °C). Labelled proteins in the supernatant were purified by affinity purification over column containing NeutrAvidin agarose resin (Pierce Thermo Scientific) following the manufacture instructions with some modifications. Briefly, columns

were equilibrated with wash buffer (10 mM PBS, pH 8, 1% NP-40) and the exactly same volume of each protein samples as well the control was load and kept in contact with the resin at room temperature for 15 min. Unlabelled proteins were washed away by 10 column volumes with wash buffer. Finally, labelled proteins were eluted (2 % SDS, 20 % glycerol, 62.5 mM Tris-HCl, 50 mM DTT, 5 % β-mercaptoethanol).

Isolation of outer membrane vesicules and precipitation of extracellular proteins.

Supernatant was filtered through a 0.45 µm vacuum filter, and the filtrate was concentrated using ultrafiltration chamber with Biomax 500 kDa ultrafiltration discs. The concentrated supernatant containing outer membrane vesicles (OMVs) was prepared by pelleting at 120 000 g for 3 hours at 4°C. The purified OMVs were diluted in phosphate-buffered saline (PBS).

Ultrafiltred supernatant was added with 0.2 mM phenylmethylsulfonyl fluoride to inhibit protease activity. Supernatant was concentrated in a final volume of about 5 mL using a Vivacell-100 centrifugal concentrator (5 kDa cut-off; Sartorius Stedim). Sodium deoxycholate (0.2 mg mL^{-1}) was added to the solution before incubation for 30 min on ice. Sodium deoxycholate supports protein precipitation, which was carried out by adding 10% (w/v) trichloroacetic acid with incubation overnight at 4 °C. After centrifugation (12,000 rpm, 30 min, 4 °C), the precipitate was washed with ice-cold acetone and solubilised in Laemmli buffer.

SDS PAGE and protein preparation for LC-MS/MS

Surface, extracellular and OMVs proteins were prepared for mass spectrometry analysis. To eliminate interfering molecules (such as salts) and concentrate proteins in a single band of polyacrylamide gel, samples were submitted to a short SDS-PAGE migration (T=12.5%, C=3.3%) in a BioRad Mini Protean II unit. Protein samples were just allowed to concentrate at the junction between stacking and resolving gel after few minutes migration (25 mA/gel at 100

V). Gel was then stained with Coomassie Brilliant Blue G250 and the single bands in each lane were excised. The proteins in the gel were reduced in 100 mM ammonium bicarbonate with 45 mM dithiothreitol for 45 min at 50 °C, and alkylated with 100 mM iodoacetamide for 20 min at room temperature in the dark. The gel was washed in 25 mM ammonium bicarbonate in 5% ACN for 30 min and twice in 25 mM ammonium bicarbonate in 50% ACN for 30 min each. The gel was dehydrated with 100% acetonitrile To stop the tryptic digestion and extract the proteins from the gel, 0.1% TFA in 100% ACN was added. Finally, samples were hydrolysed overnight with 600 ng of trypsin (Promega, Trypsine) and the peptides extracted with 100 % acetonitrile. Hydrolysed samples were purified (technical duplicates) by nanoHPLC (Ultimate 3000, Dionex) with a concentration column (500 µm/5 mm). Based on the hydrophobicity, peptides were separated with a C18 column (15 cm/1 mm) with acetonitrile solution gradient (80/20 ACN/H_2O, 0.5 % Formic acid) from 4 to 90 % during 50 min. NanoHPLC is coupled via nanoCaptiveSpray (BRUKER) to Ultra-High Resolution Qq-Time-Of-Flight (UHR QqTOF, IMPACTII, Bruker) in collision-induced dissociation mode (ProteinID-InstantExpert_IBOn.m-profile).

Data processing and bioinformatic analysis

The proteins identification was performed from *E. coli* O157:H7 strain EDL933 database (T number T00044) using ProteinScape (Bruker v1.1). Peptides were validated for a Mascot percolator score permitting to obtain a false discovery rate (FDR) below at 1 % and proteins were identified when a single high-quality tandem MS spectrum of a peptide matched significantly with the database and validated for a Mascot score higher than 17. Relative quantitation of proteins was performed following the standard workflow provided by Progenesis QI for proteomics (nonlinear dynamics, Waters). Differential protein quantification was considered significant for fold abundance greater than 1.5 and *p*-value lower than 0.05.

Post-identification analysis was performed using different software tools to predict subcellular localisations, biological processes and pathway associations. *In silico* predictions of subcellular protein localisations were accessed using PSORTb v.3.0 (www.psort.org/psortb/) combined with SignalP (export signal peptides) and SecretomeP (non-classically secreted proteins), online at www.cbs.dtu.dk/. The enrichment of gene ontology (GO) for molecular function was performed using STRAP (Software for Researching Annotations of Proteins) [33], which retrieved overrepresented GO terms related with molecular function based on UniprotKB and EBI QuickGO. Identified proteins predicted as located at the cytoplasm were filtered for moonlight activity using a moonlight proteins database (http://www.moonlightingproteins.org). Network analysis was set on STRING database v9.0 (http://www.string-db.org/). Proteins were linked based on seven criteria: neighbourhood, gene fusion, cooccurrence, coexpression, experimental evidences, existing databases, and text mining.

RESULTS

Proteins differencial abundant profiling on extracellular, surface and released OMVs sub proteomes of *E. coli* O157:H7 str. EDL933.

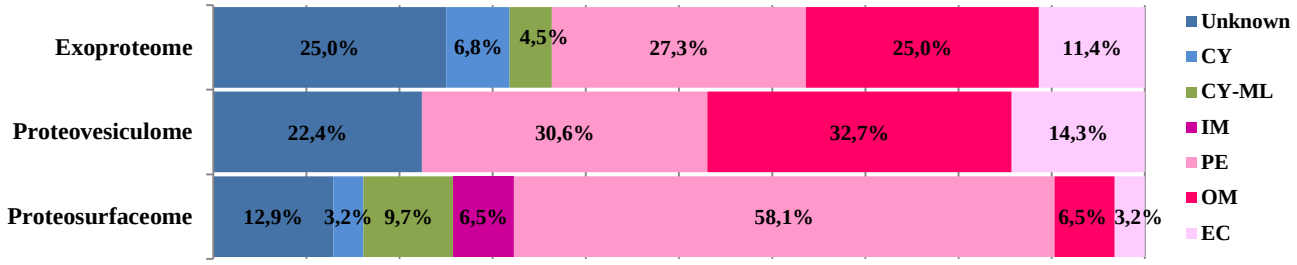
To evaluate the influence of the different cultrure media in the secreted proteins on three specific sub proteome (surface, outer membrane vesicles and extracellular) of pathogenic enterohaemonrragic *E. coli* O157:H7 str. EDL933, a label-free quantitative proteomic approach was applied. A total of 169 unique proteins were identified, corresponding to 57, 127 and 58 proteins for surface, extracellular and OMVs sub proteomes, respectivly (Table 1S). Protein abundance shifting were analysed individually for each sub proteome comparing all the three culture media. Comparative analysis revealed significantly differences in proteins abundace

depending on the culture medium. Proteins showing differential abundance in the three culture conditions were further referred as differential proteins (Tables 1, 2 and 3). Regarding the proteosurfaceome (Table 1), 31 out of 57 proteins, corresponding to 54% of total identified proteins in this subproteome, show a change in their abundace: in DMEM, 22 out of the 31 differentiral proteins presented the highest abundance; in the contrary, BHI was the condition where more proteins present the lowest abundance (24 out of 31). Concerning the proteovesiculome (Table 2), 49 out of 127 proteins (39%) have differential abundance. In M9, 38 out of those 49 proteins were highest abundant comparing to the other two media where in DMEM the lowest abundance of differential proteis was observed (21 out of 49 proteins). The exoproteome (Table 3) was the sub proteome in which the most important variability in protein abundance was registered. In total, 76% of the proteins identified, *i.e.* 44 out of 58 proteins, were differentially abundant. For the exoproteome, M9 was the culture condition with more proteins showing the highest abundance (26 out of 44), contrasting with BHI where 19 out of 58 proteins showed the lowest abundance.

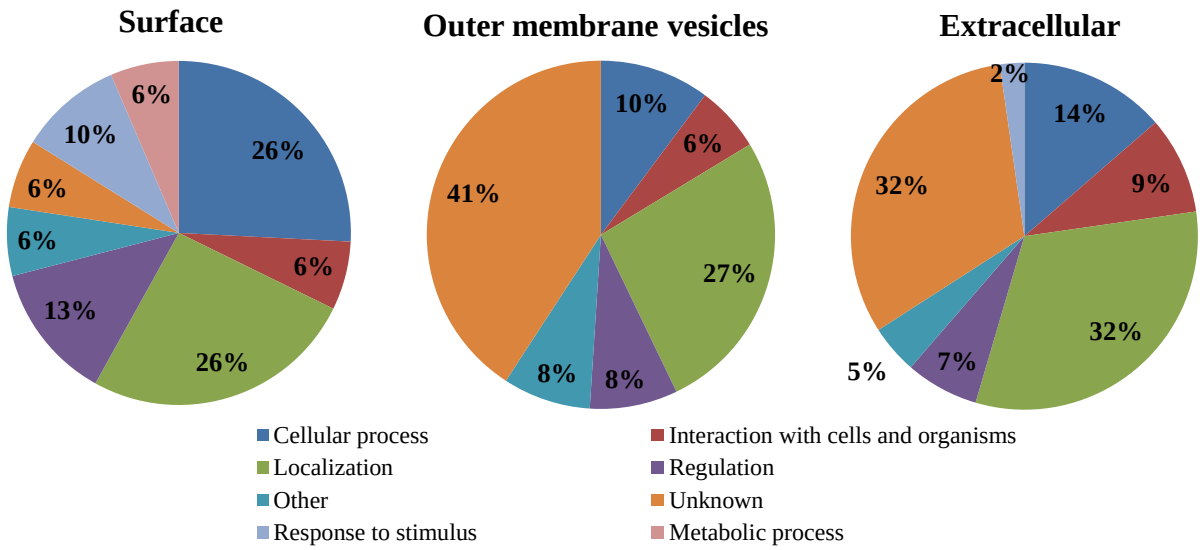
In order to identify which biological processes were afected by these different culture conditions and in these different subproteomes, the differential proteins were categorised according to the gene ontology (GO) classes and the subcellular localisation (SCL) prediction (Figure 1). In terms of biological processes, the distribution of categories for the differential proteins identified in proteovesiculome and exoprotoeme were quite similar; localisation was the process with the most proteins were associated with (27% and 32% respectivly). Other processes presented very close percentages, such cellular process, regulation and interaction with cells and organisms. Comparing to the distribution of the differential proteins identified in the proteosurfaceome, only localisation presented similar percentage of proteins associated (26%); at this subproteome,

cellular process, regulation and response to stimulus were the prominent biological processes associated with the differential proteins. Also the percentage of differential proteins with non assigned function (Unknown) was higher in the proteovesiculome and exoproteome than in the proteosurfaceome. Concerning the SCL prediction, the exoproteome and proteovesiculome presented again similar profiles but differing from protoesurfaceome. Proteins abundances changes were more perceptible on proteins predicted as located in the periplasm (GO: 0042597) and at the outer membrane (OM, GO: 0036406) with both SCL representing more than 50% of all differential proteins. Proteins predicted as extracellular (GO:0005576) represented more than 10% in both the exoproteome and protoevesiculome (Figure 1A). For the proteosurfaceome, proteins predicted as located in the periplasm represented 58.1% of the differential proteins, whereas both exoproteins (extracellular proteins) and OMPs (outer membrane proteins) represented less than 10%. In the proteosurfaceome, proteins predicted as located at the inner membrane (IM, GO: 0005886) represented 6.5% of the differntial proteins. Interestingly, some proteins predicted as cytoplasmic (GO: 0005737) were already charaterised as moonlight proteins; in the proteosurfaceome and exoproteome, these cytoproteins represented 9.7% and 4.5% of the differential proteins respectively (Figure 1A).

A



B



Biological Process (GO:0008150)

GO subcategory	GO sub-subcategory	Associated GO Term
Cellular Process		GO:0009987
Interaction with cells and organisms		Custom Category
	Biological Adhesion	GO:0022610
	Multi-organism Process	GO:0051704
	Multicellular Organismal Process	GO:0032501
Localisation		GO:0051179
Metabolic Process		GO:0008152
Regulation		Custom Category
	Regulation of biological process	GO:0050789
	Negative regulation of biological process	GO:0048519
	Positive regulation of biological process	GO:0048518
Response to stimulus		GO:0050896

Figure 1. (A) Distribution of predicted subcellular location among differential proteins identified in cell surface, OMVs and extracellular sub proteomes. Pink shades are related to proteins predicted as located within the cell envelope (GO: 0030313), i.e. either at the outer membrane (OM; GO: 0036406), in the periplasm (PE; GO: 0042597), or at the inner membrane (IM; GO: 0005886). Blue colours stands for proteins predicted as sublacted in the cytoplasm (CY; GO: 0005737) or with unknown localisation. Green stands for proteins predicted as cytoplasm but with surface moonlight (CY-ML) activity described.
(B) Distribution of the biological process terms according to the gene ontology (GO) among the differential proteins identified in cell surface, OMVs and extracellular sub proteomes. Description of Go terms for each individual biological process.

Secreted proteins associated with secretion sytems are influenced by the culture medium

Once characterisation of the biological processes revealed Localisation as the process with more proteins altering their abundance, the number of proteins higher or lower abundant per biological process were analysed (Figure 2). This analsys revealed that, M9 was influencing positivly the abundance of proteins related with Localisation to DMEM or BHI which the majority of the proteins here quantified showed the lowest abundance. The relative abundance between single proteins can be observed in figure 3. Following the heat maps is possible to note that the great majority of quantified proteins on the proteovesiculome and exoproteome presented the highest abundance in M9, however at the level of proteosurfaceome DMEM was the condition in which quantified proteins with highest abundance were the majority. Figure 4 was, then designed merging the information of the heat map and a standard STRING analysis.

STRING network analysis was performed with the purpose to correlate proteins in functional clusters. Considering localization process was identified as the most affected in this study, the

investigation of proteins involved with mechanisms of secretion was addressed. STRING predicted the functional connections between some secreted proteins and their cognate protein secretion systems and with the relative quantitation from heat maps the highest abundant proteins were highlighted. Proteins related to the T3aSS showed highest abundance when bacteria were cultured in DMEM (red shade, Figure 4). Particularly, in the exoproteome and protoevesiculome (Figure 3 and Figure 4), the proteins associated with T3SS structure, e.g. EspA, EspB and EspD, were up to 50 fold higher in abundance in DMEM compared to BHI or M9. In the proteosufaceome (Figure 4C), intimin (encoded by *eae*) showed a 10-fold higher abundance in DMEM. The chaperone proteins DegP (*ptd*) and Skp were higher abundant in DMEM and this condition also reveals the highest abundance in a group of moonlighting ribosomal subunits (RplC, RplS and RplA) as well for StcE, a metallo proteins secreted by the T2SS. In M9 condition (grey shade, Figure 4), several clusters of proteins higher abundant in this condition were revealed. In the proteovesiculome and exoproteome, the first cluster includes several porins, e.g. OmpA/C/F/X/T, which were up to 10 fold more abundant in M9 compared to DMEM or BHI. Another clearly defined cluster was related to differential proteins related to the T1SS, namley LivJ, LivK, FliY, GlnH and YbeJ binding proteins. Proteins fom this cluster were higher abundance in M9 condition in the exoproteome and proteovesiculome. In BHI (green shade, Figure 4), the higher abundant proteins are related with the T3bSS, namely some flagellar proteins, some of being also identified in the proteovesiculome. FlgG, FlgL, FliC, FliC (*flbC*) and FlgK were up to 40 fold more abundant in BHI than in DMEM or M9 (Figure 4B).

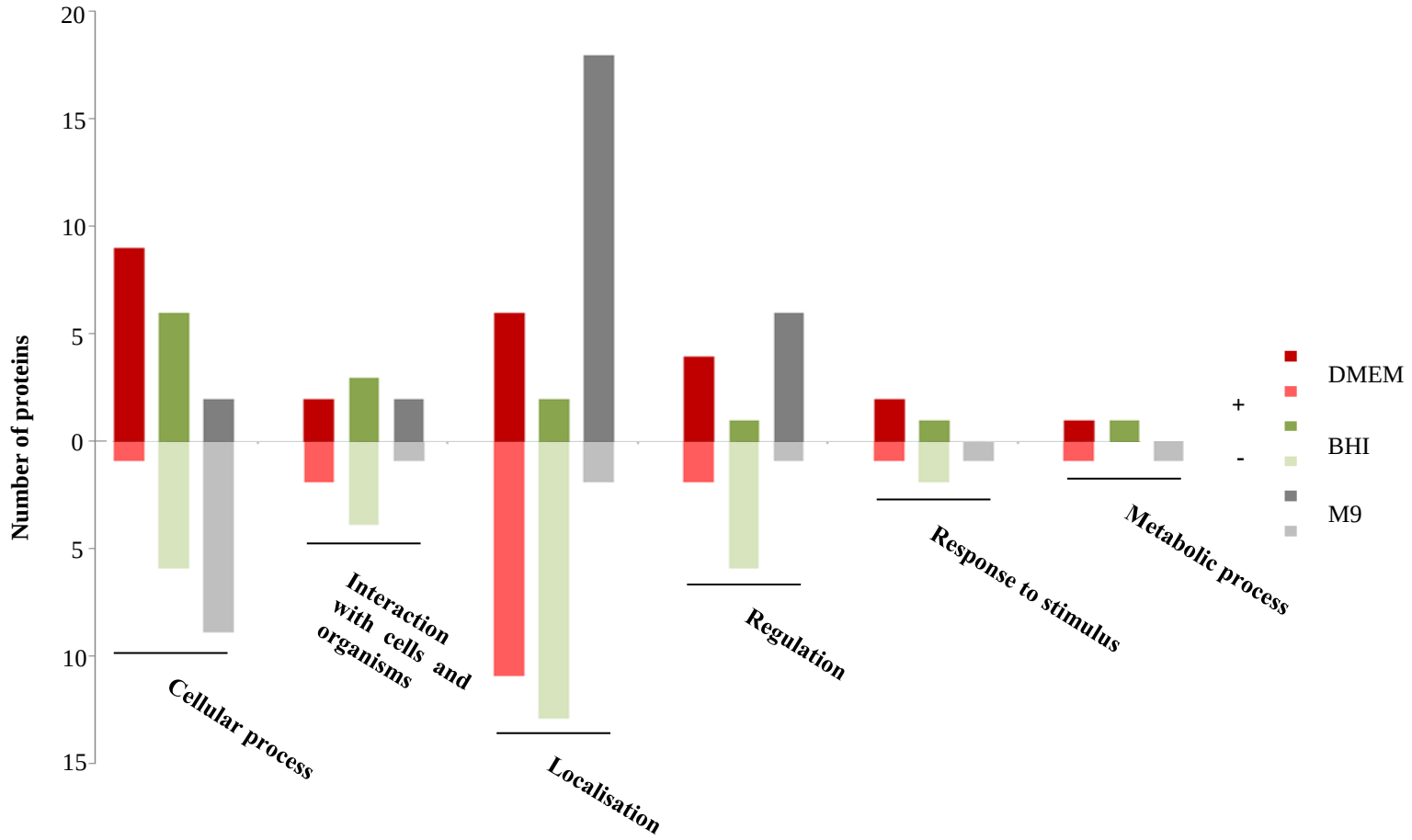


Figure 2. Biological processes related to differential protein abundance depending on the culture medium condition. Bars above 0 represents proteins higher abundant and under 0 proteins lower abundant. Red shades stand for DMEM, green for BHI and grey for M9.

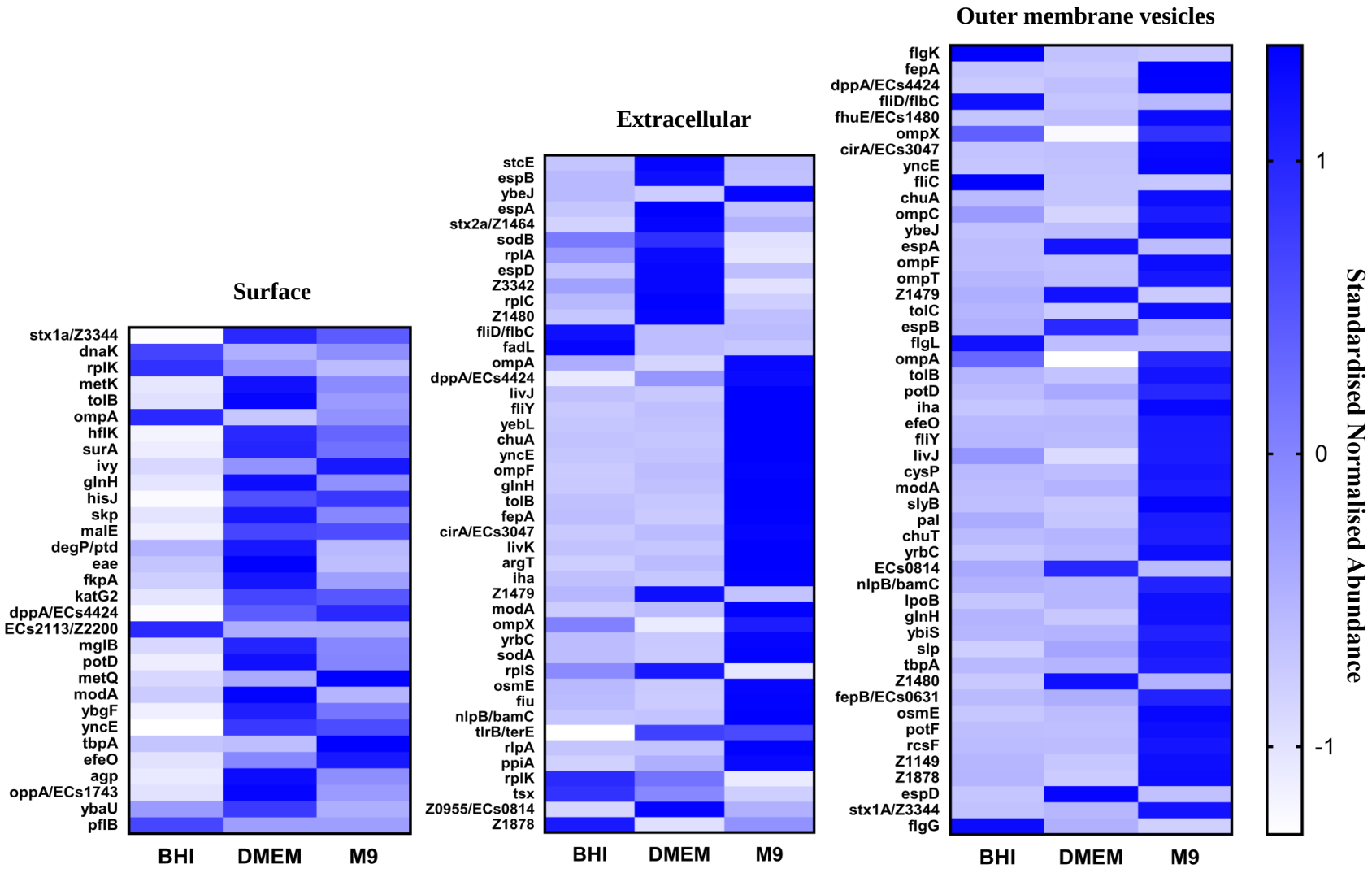
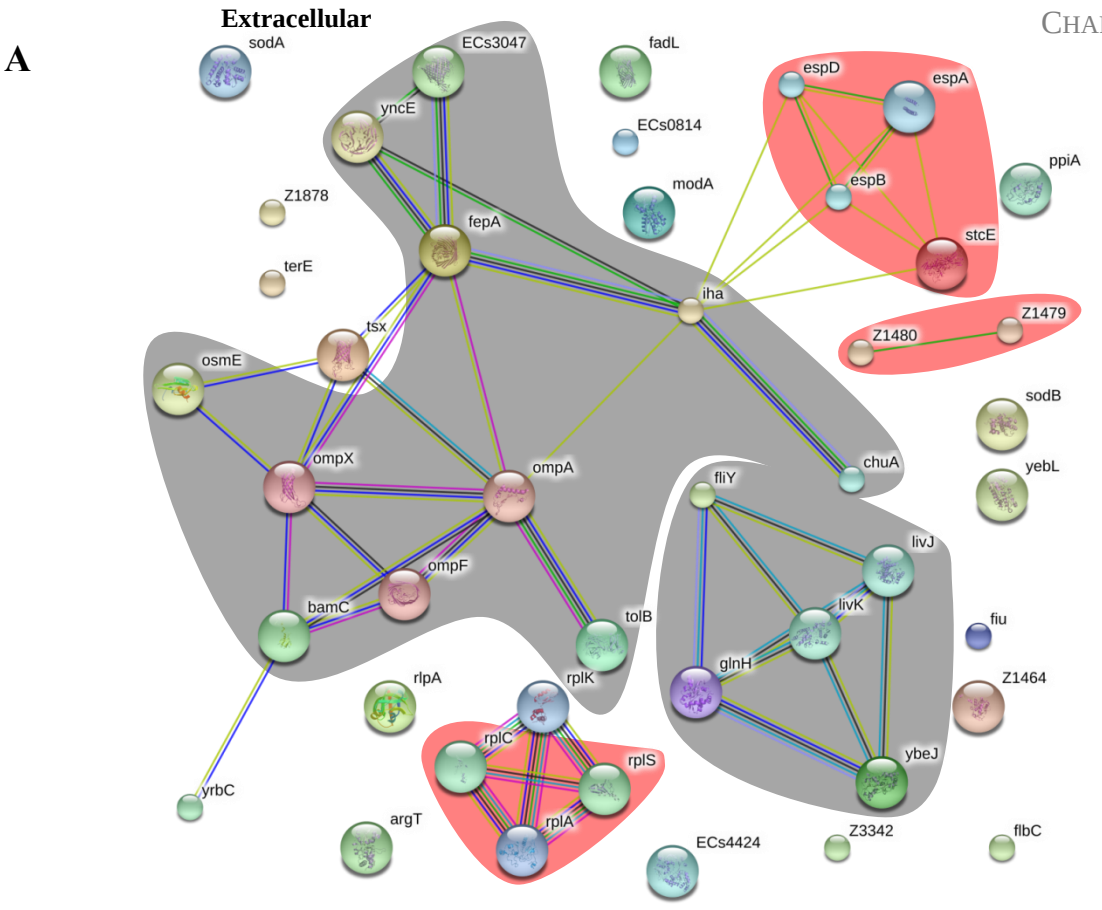
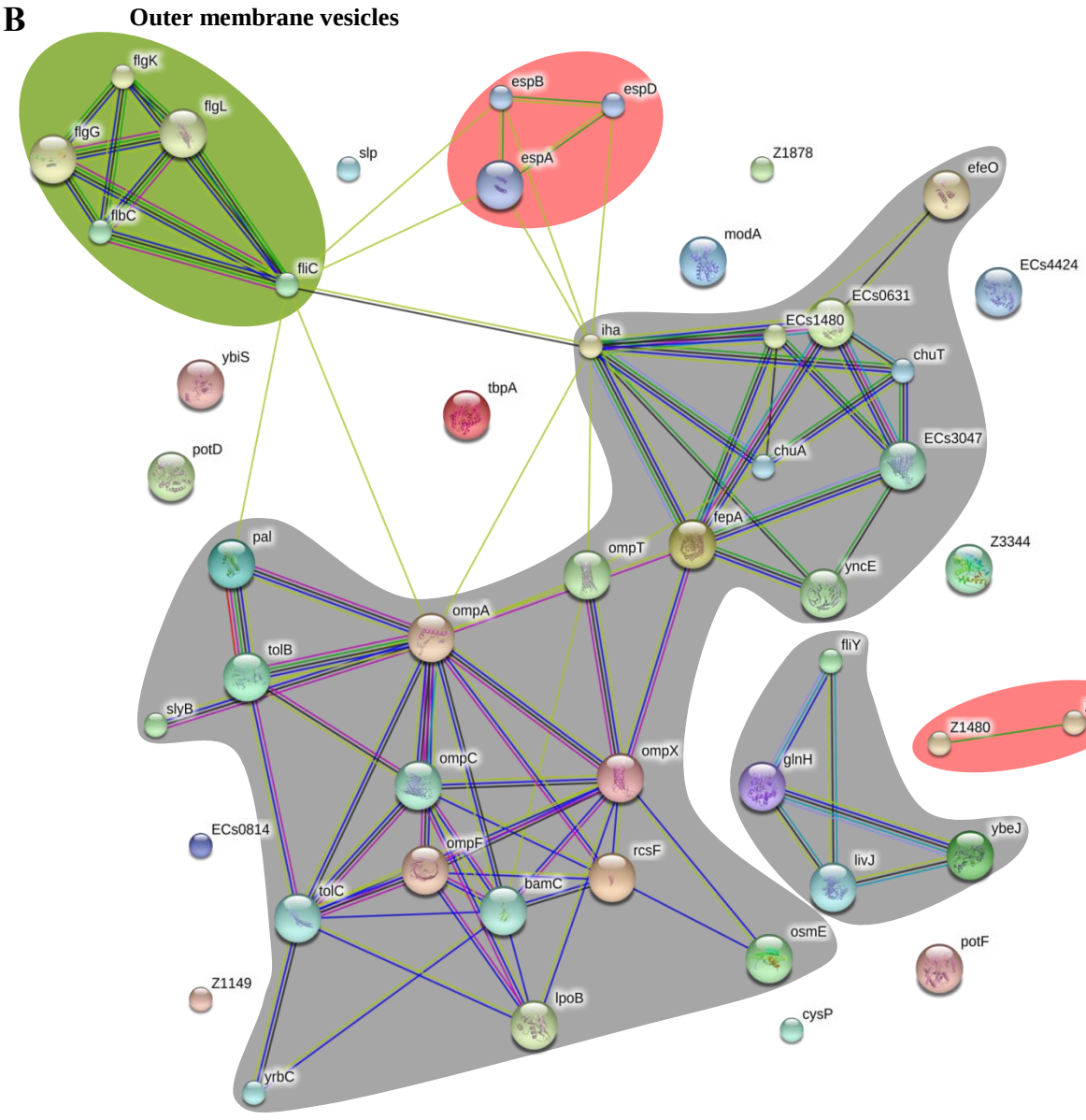


Figure 3. Heat maps of standardised normalised abundance of proteins significantly different in surface, extracellular and OMVs sub proteomes and DMEM, BHI and M9 media. 31, 44, and 49 were compared for each sub proteome respectivly. Blue colour intensity represents the relative abundace between proteins (dark blue stands for higher abundace).

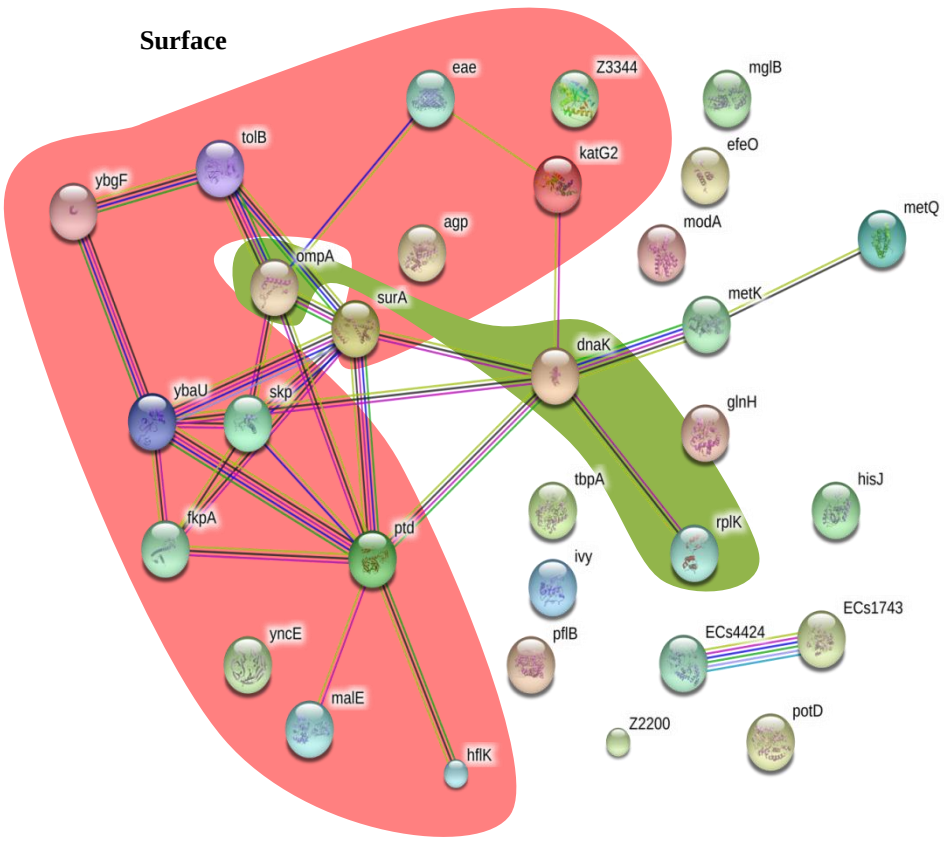
A **Extracellular**



B **Outer membrane vesicles**



C **Surface**



Nodes:
Network nodes represent proteins
splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Size

small nodes:
protein of unknown 3D structure

large nodes:
some 3D structure is known or predicted

Node Color

colored nodes:
query proteins and first shell of interactors

white nodes:
second shell of interactors

Edges:
Edges represent protein-protein associations
associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.

Known Interactions

from curated databases
experimentally determined

Predicted Interactions

gene neighborhood
gene fusions
gene co-occurrence

Others

textmining
co-expression
protein homology

Higher abundant proteins

BHI DMEM M9

Figure 4. Protein–protein interaction network of differential abundant proteins obtained with STRING v. 9.1. Colour shades mark proteins higher abundant; red stands for DMEM, green for BHI and grey for M9 culture medium. A, B and C sections represents the three sub proteomes, extracellular, OMVs and surface, respectivly.

DISCUSSION

Numerous investigations reported differential expression of virulence factors, including adhesins, depending on environmental conditions. However these studies were focused mainly on transcriptional regulation [16-19, 26-31] or on the adhesion phenotype to *in vitro* models [14, 18, 32, 33] and/or *in vivo* models [14]. Transcriptome analyses have provided high level of gene regulation, however these data is not extended to protein turnover, and therefore do not provide a complete panorama of the proteins expression [34]. The purpose behind this study was that *E. coli* O157:H7 strain EDL933 secretome should largely reflect the environmental conditions in which the strain is growing. Following this premise, the influence of three culture media on EHEC O157:H7 secretome was evaluated. Culture conditions were choosen based on distintive characteristics: i) BHI is commonly used and reported as inducer of specific adhesion properties on EHEC [35]; ii) DMEM is reported as the ideal medium to express LEE *locus* [16, 17]; and iii) the minimal chemical medium M9 for culture *E. coli*, which is commonly used in bacterial physiology to allow to full expression and use of catabolic and anabolic pathways [36]. Also to get a fullpictures of the expression landscape of secreted proteins, the proteosurfaceome, exoproteome and proteovesiculome were isolated for each culture condition. Proteomic quantitative analysis revealed significant changes in all subproteomes. Particullary, proteins secreted into the extracellular medium changed their abundance up to 76% of the total identified

proteins in this sub proteome. Interestingly, minimal culture condition such M9 was the condition with more proteins highly abundant, distinctly from previous studies where M9 were associated with secretion inhibition [16]. Also the fact that the biological process Localisation (GO:0051179) was identified as the biological process most affected by culture conditions, supports the evidence of active proteins transport and movement in the cell.

As secretion is intrisically dependent from secretion systems complexes, our study went deeper in the investigation of those cellular structures, since in pathogenic *E. coli*, secretion systems are highly involved in the main virulence features. Here we highlight the differential expression of proteins associated with T3SS on EHEC depending on growth medium. We reported that a subset of affected proteins by growth conditions, secreted into extracellular environment and/or through OMVs, were involved in both injectisome and flagellar filament, namly EspA/B/D and FliD (*flbC*), FliC, FlgG, FlgK, FlgL respectivly. T3aSS proteins were higher expressed when EHEC were grown in DMEM than in BHI culture medium, conditions which are associated with higher expression abundance of T3bSS proteins (Figure 3). The higher expression of proteins related with T3aSS in DMEM is not unexpected since this culture condition was reported as the optimal condition to express the LEE (encoding the T3aSS) [16, 17]. However the lower expression of T3aSS and the higher expression of T3bSS could suggest their counter-expression. Despite this was never reported in EHEC, the counter-regulation of these two types of T3SS has been observed in other pathogens such *Yersinia* and *Shigella*, but flagellar genes are often activated only at temperatures below 30°C, while injectisome genes are transcribed within the host organism, which for animal pathogens is sensed by a temperature of 37°C [37-40]. Since all our bacterial cultures were set up at 37 °C, other influent factors to the differential expression in EHEC might be involved. Further hypothesis could be the composition of the culture media. It is

known that among the components of DMEM, only sodium bicarbonate was found to remarkably stimulate bacterial adherence by enhancing the transcription of *ler*, required for LEE-encoded gene expression [26]. Indeed EspA/B as well intimin, higher expressed at the cell surface when EHEC were grown in DMEM in the present study, are positivly enhanced by sodium bicarbonate. Even though the presence of sodium bicarbonate explains T3aSS expressed in DMEM, the counter-expression of T3bSS proteins by BHI condition is still not clear. Another aspects to consider is the growth rate, which is higher in BHI than DMEM and M9 media. In *E. coli*, it is known that flagellar biosynthesis is correlated with the bacterial growth rate [41]. Furthermore investigation in that directionwould be require to determine its impact on the secretome of *E. coli*.

In M9 conditions, an increase on the expression of T1SS proteins or porins was observed. Nutrient availability is a major factor in bacterial survival and proliferation and as introduced before reduced growth rate resulting from nutrient limitation leads to regulation changes in bacterial outer-membrane composition and consequent protein secretion [42]. Porin proteins control the permeability of the outer membrane and nutrient limitation strongly and differentially regulates porin expression. One of the main differences between M9 and BHI and DMEM is the low abundance of iron ions in M9, actually under iron deprivation, there are reports about the upregulation of proteins involved in channel forming proteins, such FepA [43], and other channel proteins, such as OmpX in *E.coli* [44]. Present results of high expression of outer membrane protein OmpA, OmpX and OmpC in adverse conditions such nutrient limitation in M9 indicates the need of bacteria to preserve the cell integrity and in specific the integrity of the outer membrane.

From these data, it clearly appears that EHEC adapts its secretion systems profile depending on the environmental conditions. *E. coli* O157 dramatically modulate the levels of protein expression in the exoproteome, proteevesiculome and protesurfaceome. In M9, it must be considered the bacterial cells need to express all their biosynthetic pathways and thus their full potential to sustain bacterial growth; this condition can be regarded as to environmental conditions with nutrient limitations. In contrast, BHI or DMEM culture conditions are plenty of nutrients (nutrient blocks or macromolecules needing to be degraded) where bacteria seems to increase the abundance of specific virulence factors and motility structures. Our differential methods revealed that EHEC change the secretion pathways by which proteins are secreted regarding the surrounding conditions. Our results also pinpointed the differential expression of T3SS, an important mechanism in EHEC. T3SS looks to be counter-regulated by the composition of BHI and DMEM, increasing or decreasing the abundance of proteins associated with flagella or injectisome. The proteomic approach applied here allowed to better correlate the genomic and transcriptomic data with the bacterial phenotype, regarding EHEC gene regulation when it is growing in different environmental conditions. Despite the complexity, further investigations are needed using culture media simulating the human intestinal conditions in order to evaluate the genuine proteome changes in EHEC.

Table 1. Proteins identified from cell surface proteome fraction with differential abundance (p>0.05) in DMEM/BHI/M9.

Uniprot ID	Gene Name	Annotation	SCL [a] prediction [a]	Fold change [b]	*p*-value	+ *Condtition*	- *Condition*
Subunits of cell surface organelles							
Q8X503	ECs2113 Z2200	Putative major fimbrial subunit	EC	7.3	1.32×10^{-05}	BHI	DMEM
Outer membrane proteins (OMPs)							
P43261	*eae*	Intimin	OM	10.3	2.26×10^{-07}	M9	BHI
P0A911	*ompA*	Outer membrane protein A	OM	3.6	1.30×10^{-04}	BHI	DMEM
Periplasmic proteins							
Q8X641	*mglB*	Galactose-binding transport protein	PE	28.7	7.52×10^{-06}	DMEM	BHI
Q8XA04	*tbpA*	Thiamin-binding periplasmic protein	PE	21.0	9.10×10^{-06}	M9	BHI
Q8XDA4	*oppA* ECs1743	Oligopeptide transport periplasmic binding protein	PE	18.4	6.65×10^{-12}	DMEM	BHI
Q8XAS6	*efeO*	Iron uptake system component EfeO	PE	11.0	2.93×10^{-02}	M9	BHI
Q8XBZ6	*agp*	Periplasmic glucose-1-phosphatase	PE	7.3	1.05×10^{-05}	DMEM	BHI
Q8X8D9	*potD*	Putrescine-binding periplasmic protein	PE	7.1	6.88×10^{-07}	DMEM	BHI
P0A857	*tolB*	Protein TolB	PE	7.0	2.02×10^{-08}	DMEM	BHI
Q8X964	*ybgF*	Uncharacterised protein	PE	6.6	2.71×10^{-07}	DMEM	BHI
P0AEQ5	*glnH*	Glutamine-binding periplasmic protein	PE	5.8	5.63×10^{-05}	DMEM	BHI
Q8X4K7	*dppA* ECs4424	Dipeptide transport protein	PE	5.4	6.97×10^{-07}	DMEM	BHI
Q8X931	*modA*	Molybdate-binding periplasmic permease protein	PE	5.2	3.79×10^{-06}	DMEM	BHI
P0AEU2	*hisJ*	Histidine-binding periplasmic protein	PE	4.0	4.19×10^{-05}	M9	BHI
Q8X8V9	*metQ*	D-methionine-binding lipoprotein MetQ	IM	3.9	1.56×10^{-06}	M9	BHI
P0AEU9	*skp*	Chaperone protein Skp	PE	3.7	2.60×10^{-05}	DMEM	BHI
P0ABZ8	*surA*	Chaperone SurA	PE	2.8	1.49×10^{-04}	DMEM	BHI
P65765	*fkpA*	FKBP-type peptidyl-prolyl cis-trans isomerase FkpA	PE	2.4	3.19×10^{-04}	DMEM	BHI
P0AEY0	*malE*	Maltose-binding periplasmic protein	PE	2.2	8.43×10^{-04}	DMEM	BHI
P0AD60	*ivy*	Inhibitor of vertebrate lysozyme	PE	1.9	2.93×10^{-02}	M9	BHI
P0C0V1	*degP ptd*	Periplasmic serine endoprotease	PE	1.9	4.11×10^{-03}	DMEM	M9
Inner membrane proteins							
Q8XE55	*ybaU*	Peptidylprolyl isomerase	IM	1.5	2.32×10^{-02}	DMEM	M9
Cell surface moonlight proteins							
P0A819	*metK*	50S ribosomal protein L1	CY-ML	3.4	1.24×10^{-06}	DMEM	BHI
P0A6Z0	*dnaK*	Chaperone protein DnaK	CY-ML	2.9	5.34×10^{-03}	BHI	M9
P0A7J9	*rplK*	50S ribosomal protein L11	CY-ML	2.7	2.37×10^{-02}	BHI	M9

Cytoplasmatic proteins							
Q8XEB4	*pflB*	Formate acetyltransferase 1	CY	3.1	8.47×10^{-05}	BHI	DMEM
Non-predicted location							
Q8X9X2	*yncE*	Uncharacterised protein YncE	Unknown	18.8	8.45×10^{-11}	DMEM	BHI
A8B1H9	*stx1a* Z3344	Shiga toxin 1 subunit A	Unknown	6.2	2.20×10^{-06}	DMEM	BHI
Q7BSW8	katG2	Catalase-peroxidase 2	Unknown	3.3	6.82×10^{-05}	DMEM	BHI
P0ABC8	*hflK*	Protein HflK	Unknown	3.2	7.53×10^{-04}	DMEM	BHI

Table 2. Proteins identified from outer membrane vesicle proteome fraction with differential abundance (p>0.05) in DMEM/BHI/M9.

Uniprot ID	Gene Name	Annotation	SCL [a] prediction [a]	Fold change [b]	*p*-value	+ *Condtition*	- *Condition*
Subunits of cell surface organelles							
Q7DB81	*espD*	EspD	EC	50.9	4.89×10^{-13}	DMEM	BHI
P58297	*fliD flbC*	Flagellar hook-associated protein 2	EC	40.3	1.14×10^{-11}	BHI	DMEM
Q7DBI0	*fliC*	Flagellin	EC	39.1	1.46×10^{-09}	BHI	M9
Q8X8K5	*flgK*	Flagellar hook-associated protein 1	EC	26.1	3.87×10^{-14}	BHI	M9
P0ABX7	*flgG*	Flagellar basal-body rod protein FlgG	EC	21.4	7.86×10^{-09}	BHI	M9
Q8X9U3	Z1149	Uncharacterised protein	EC	8.4	1.92×10^{-08}	M9	DMEM
Q8X8K3	*flgL*	Flagellar biosynthesis hook-filament junction protein	EC	3.9	5.52×10^{-04}	BHI	M9
Outer membrane proteins (OMPs)							
P0A919	*ompX*	Outer membrane protein X	OM	9.8	6.52×10^{-11}	M9	DMEM
Q8XBW7	*fepA*	Outer membrane receptor for ferric enterobactin (Enterochelin) and colicins B and D	OM	9.8	3.87×10^{-14}	M9	DMEM
P0A911	*ompA*	Outer membrane protein A	OM	8.2	4.41×10^{-13}	M9	DMEM
Q9LAP1	*iha*	Adhesin	OM	8.2	1.89×10^{-09}	M9	BHI
Q8X8H4	*fhuE* ECs1480	Outer membrane receptor for ferric iron uptake	OM	8.7	5.96×10^{-11}	M9	BHI
Q8XE41	*ompC*	Outer membrane protein C	OM	8.5	3.55×10^{-08}	M9	DMEM
Q7DB97	*chuA*	Heme utilisation/transport protein	OM	7.9	9.31×10^{-09}	M9	DMEM
Q8X633	*cirA* ECs3047	Outer membrane receptor for iron-regulated colicin I receptor porin requires tonB gene product	OM	7.4	2.67×10^{-10}	M9	BHI
Q8XDF1	*ompF*	Outer membrane protein 1a	OM	6.5	1.31×10^{-07}	M9	DMEM
P58603	*ompT*	Protease 7	OM	4.1	1.55×10^{-06}	M9	DMEM
Q8XBP7	*tolC*	Outer membrane channel TolC	OM	3.0	1.04×10^{-04}	M9	DMEM

Outer membrane lipoproteins (OMLs)							
P0A913	*pal*	Peptidoglycan-associated lipoprotein	OM	11.1	6.72×10^{-06}	M9	DMEM
P0A906	*slyB*	Outer membrane lipoprotein SlyB	OM	9.6	1.31×10^{-08}	M9	DMEM
Q8XBD3	*nlpB bamC*	Outer membrane protein assembly factor BamC	OM	7.9	2.95×10^{-02}	M9	DMEM
Q8X8W0	*rcsF*	Outer membrane lipoprotein RcsF	OM	6.5	6.31×10^{-07}	M9	DMEM
Q8X5P5	*slp*	Outer membrane protein Slp	OM	2.8	1.76×10^{-04}	M9	BHI
Periplasmic proteins							
Q8XBV4	*fepB* ECs0631	Ferric enterobactin (Enterochelin) binding protein periplasmic component	PE	25.2	2.22×10^{-03}	M9	BHI
Q8XBL6	*ybeJ*	Putative periplasmic binding transport protein	PE	15.0	5.62×10^{-08}	M9	BHI
P0AD98	*livJ*	Leu/Ile/Val-binding protein	PE	12.4	6.63×10^{-07}	M9	DMEM
Q8X4K7	*dppA* ECs4424	Dipeptide transport protein	PE	11.6	3.62×10^{-13}	M9	BHI
Q8XA04	*tbpA*	Thiamin-binding periplasmic protein	PE	9.3	2.91×10^{-04}	M9	BHI
Q8XBJ5	*cysP*	Thiosulfate binding protein	PE	8.8	2.05×10^{-06}	M9	DMEM
P0A857	*tolB*	Protein TolB	PE	8.6	8.65×10^{-07}	M9	DMEM
Q8X931	*modA*	Molybdate-binding periplasmic permease protein	PE	8.6	1.97×10^{-04}	M9	BHI
Q8XAS6	*efeO*	Iron uptake system component EfeO	PE	8.5	1.77×10^{-05}	M9	BHI
Q8XC86	*espB*	EspB	PE	8.0	1.08×10^{-04}	DMEM	M9
P0AEQ5	*glnH*	Glutamine-binding periplasmic protein	PE	7.8	4.40×10^{-07}	M9	DMEM
Q8XBC5	*fliY*	Putative periplasmic binding transport protein	PE	7.2	9.30×10^{-05}	M9	DMEM
Q8X5N7	*chuT*	Putative hemin binding protein	PE	7.0	5.93×10^{-04}	M9	BHI
Q8X6R4	*potF*	Putrescine-binding periplasmic protein	PE	6.6	4.12×10^{-08}	M9	DMEM
Q8X8D9	*potD*	Putrescine-binding periplasmic protein	PE	3.2	1.76×10^{-02}	M9	BHI
Non-predicted location							
Q8X9X2	*yncE*	Uncharacterised protein YncE	Unknown	42.1	5.03×10^{-10}	M9	BHI
Q7DB80	*espA*	EspA	Unknown	27.5	1.19×10^{-07}	DMEM	M9
P0ADB2	*osmE*	Osmotically-inducible lipoprotein E	Unknown	25.8	4.01×10^{-10}	M9	BHI
P0AB39	*lpoB*	Penicillin-binding protein activator LpoB	Unknown	25.7	1.65×10^{-09}	M9	BHI
P0AAY0	*ybiS*	Probable L.D-transpeptidase YbiS	Unknown	11.6	3.53×10^{-03}	M9	BHI
Q8X9J7	*yrbC*	Uncharacterised protein	Unknown	9.6	2.19×10^{-08}	M9	BHI
Q8X4P5	Z1878	Bor protein	Unknown	9.2	7.99×10^{-08}	M9	DMEM
A8B1H9	*stx1A* Z3344	Shiga toxin 1 subunit A	Unknown	7.0	9.51×10^{-06}	M9	BHI
A7UQW6	Z1479	Uncharacterised protein	Unknown	5.9	1.49×10^{-05}	DMEM	M9
Q8X881	ECs0814	Uncharacterised protein	Unknown	4.1	3.45×10^{-03}	DMEM	M9
A7UQW5	Z1480	Uncharacterised protein	Unknown	3.5	2.73×10^{-03}	DMEM	BHI

Table 3. Proteins identified from extracellular proteome fraction with differential abundance (p>0.05) in DMEM/BHI/M9.

Uniprot ID	Gene Name	Annotation	SCL [a] prediction [a]	Fold change [b]	*p*-value	+ *Condtition*	- *Condition*
Subunits of cell surface organelles							
O82882	*stcE*	Metalloprotease StcE	EC	24.4	2.92×10^{-13}	DMEM	BHI
Q8XC86	*espB*	EspB	EC	18.8	3.06×10^{-09}	DMEM	M9
Q7DB81	*espD*	EspD	EC	18.0	5.53×10^{-11}	DMEM	BHI
P58297	*fliD flbC*	Flagellar hook-associated protein 2	EC	13.1	4.74×10^{-07}	BHI	DMEM
Q8XBQ1	*rlpA*	Rare lipoprotein A	EC	10.4	1.41×10^{-12}	M9	BHI
Outer membrane proteins (OMPs)							
Q8XCN6	*fadL*	Long-chain fatty acid transport protein	OM	21.5	1.30×10^{-12}	BHI	M9
Q8X7W7	*fiu*	Catecholate siderophore receptor Fiu	OM	19.8	6.67×10^{-10}	M9	DMEM
Q7DB97	*chuA*	Heme utilisation/transport protein	OM	11.2	8.23×10^{-13}	M9	DMEM
P0A911	*ompA*	Outer membrane protein A	OM	10.6	1.43×10^{-09}	M9	DMEM
Q8XDF1	*ompF*	Outer membrane protein 1a	OM	8.0	1.32×10^{-12}	M9	BHI
Q9LAP1	*iha*	Adhesin	OM	5.8	1.85×10^{-13}	M9	DMEM
Q8XBW7	*fepA*	Outer membrane receptor for ferric enterobactin (Enterochelin) and colicins B and D	OM	5.7	3.81×10^{-12}	M9	DMEM
P0A919	*ompX*	Outer membrane protein X	OM	4.9	5.47×10^{-08}	M9	DMEM
Q8X633	*cirA* ECs3047	Outer membrane receptor for iron-regulated colicin I receptor porin requires tonB gene product	OM	4.7	4.90×10^{-12}	M9	BHI
P0A928	*tsx*	Nucleoside-specific channel-forming protein tsx	OM	2.2	1.92×10^{-03}	BHI	M9
Outer membrane lipoproteins (OMLs)							
Q8XBD3	*nlpB bamC*	Outer membrane protein assembly factor BamC	OM	8.0	5.70×10^{-11}	M9	BHI
Periplasmic proteins							
Q8X4K7	*dppA* ECs4424	Dipeptide transport protein	PE	14.3	1.89×10^{-12}	M9	BHI
Q8XCJ1	*yebL*	Putative adhesin	PE	13.0	1.32×10^{-12}	M9	BHI
P0AGD5	*sodB*	Superoxide dismutase [Fe]	PE	10.1	2.17×10^{-05}	DMEM	M9
P0AD98	*livJ*	Leu/Ile/Val-binding protein	PE	10.1	3.00×10^{-10}	M9	DMEM
Q8XBC5	*fliY*	Putative periplasmic binding transport protein	PE	9.0	1.89×10^{-12}	M9	BHI
Q8X6S7	*livK*	High-affinity leucine-specific transport system periplasmic binding protein	PE	8.9	2.09×10^{-14}	M9	DMEM
Q8XBL6	*ybeJ*	Putative periplasmic binding transport protein	PE	8.2	7.70×10^{-08}	M9	DMEM
P0AFL5	*ppiA*	Peptidyl-prolyl cis-trans isomerase A	PE	7.4	1.95×10^{-11}	M9	BHI
Q8XCS4	*argT*	Lysine-. arginine-. ornithine-binding	PE	7.1	7.62×10^{-11}	M9	BHI

		periplasmic protein					
P0A857	*tolB*	Protein TolB	PE	5.9	1.85×10^{-13}	M9	DMEM
P0AEQ5	*glnH*	Glutamine-binding periplasmic protein	PE	4.6	5.72×10^{-13}	M9	BHI
Q8X931	*modA*	Molybdate-binding periplasmic permease protein	PE	4.1	1.48×10^{-09}	M9	BHI
Cell surface moonlight proteins							
P0A7J9	*rplK*	50S ribosomal protein L11	CY-ML	27.4	1.26×10^{-09}	BHI	M9
P0A7L2	*rplA*	50S ribosomal protein L1	CY-ML	14.7	3.04×10^{-13}	DMEM	M9
Cytoplasmatic proteins							
P60440	*rplC*	50S ribosomal protein L3	CY	49.8	2.65×10^{-11}	DMEM	M9
P0A7K8	*rplS*	50S ribosomal protein L19	CY	4.2	1.65×10^{-07}	DMEM	M9
Q9LAN9	*tlrB/terE*	Putative phage inhibition. colicin resistance and tellurite resistance protein	CY	3.6	4.31×10^{-06}	M9	BHI
Non-predicted location							
Q8X9X2	*yncE*	Uncharacterised protein YncE	Unknown	18.2	5.55×10^{-16}	M9	BHI
Q7DI68	*stx2a* Z1464	Shiga toxin 2 A-subunit	Unknown	10.1	3.48×10^{-12}	DMEM	BHI
A8B1I0	Z3342	Uncharacterised protein	Unknown	9.7	3.50×10^{-12}	DMEM	M9
Q7DB80	*espA*	EspA	Unknown	5.9	1.15×10^{-09}	DMEM	BHI
A7UQW6	Z1479	Uncharacterised protein	Unknown	4.9	1.47×10^{-06}	DMEM	M9
P66828	*sodA*	Superoxide dismutase [Mn]	Unknown	4.7	1.55×10^{-10}	M9	DMEM
Q8X881	Z0955 ECs0814	Uncharacterised protein	Unknown	4.3	2.77×10^{-08}	DMEM	BHI
P0ADB2	*osmE*	Osmotically-inducible lipoprotein E	Unknown	3.9	7.78×10^{-10}	M9	DMEM
A7UQW5	Z1480	Uncharacterised protein	Unknown	3.9	1.39×10^{-08}	DMEM	BHI
Q8X9J7	*yrbC*	Uncharacterised protein	Unknown	3.5	1.27×10^{-10}	M9	DMEM
Q8X4P5	*Z1878*	Bor protein	Unknown	2.1	4.93×10^{-04}	BHI	DMEM

[a] Predicted subcellular localisation based on PSORTb v3.0, EC: extracellular (GO: 0005576), OM: outer membrane (GO:0036406), PE: periplasm (GO: 0042597), IM: inner membrane (GO: 0005886), CY: cytoplasm (GO: 0005737), CY-ML: cytoplasmic proteins with moonlight activity.

[b] Fold change corresponds to the differential protein abundance from DMEM, BHI, M9 comparison. Fold change was mesure comparing the highest conditions with the lowest condition. Differences were considered significant for *p*-values below 0.05.

ACKNOWLEDGEMENTS

This work was supported in part by INRA (Institut National de la Recherche Agronomique), GSK and by EU FP7 Marie Curie Actions ITN (Initial Training Network) EID (European Industrial Doctorates) DIS*Co* (A multidisciplinary Doctoral Industrial School on novel preventive strategies against *E. Coli* infections; http://www.discoproject.eu/; n°FP7-PEOPLE-607611) project. Ricardo MONTEIRO is a Marie Curie PhD Research Fellow granted by ITN EID DIS*Co*.

SUPPLEMENTARY MATERIAL

Table 1S: Total proteins identified in surface, outer membrane and extracellular proteome fractions.

REFERENCES

1. Hooper LV, Gordon JI. Commensal host-bacterial relationships in the gut. *Science* 2001; **292**: 1115-8.
2. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 1998; **11**: 142-201.
3. Cohen MB, Nataro JP, Bernstein DI et al. Prevalence of diarrheagenic *Escherichia coli* in acute childhood enteritis: a prospective controlled study. *J Pediatr* 2005; **146**: 54-61.
4. Karch H, Tarr PI, Bielaszewska M. Enterohaemorrhagic *Escherichia coli* in human medicine. *Int J Med Microbiol* 2005; **295**: 405-18.
5. Ahmed M, Van Velkinburgh J. Enterohemorrhagic *Escherichia coli* O157 in North Africa region: a threat require advanced investigation. *Pan Afr Med J* 2014; **19**: 26.
6. Frankel G, Phillips AD, Rosenshine I et al. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Mol Microbiol* 1998; **30**: 911-21.
7. Elliott SJ, Wainwright LA, McDaniel TK et al. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* 1998; **28**: 1-4.
8. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* 1995; **92**: 1664-8.
9. Donnenberg MS, Tzipori S, McKee ML et al. The role of the eae gene of enterohemorrhagic *Escherichia coli* in intimate attachment in vitro and in a porcine model. *J Clin Invest* 1993; **92**: 1418-24.
10. McKee ML, Melton-Celsa AR, Moxley RA et al. Enterohemorrhagic *Escherichia coli* O157:H7 requires intimin to colonize the gnotobiotic pig intestine and to adhere to HEp-2 cells. *Infect Immun* 1995; **63**: 3739-44.
11. Tzipori S, Gibson R, Montanaro J. Nature and distribution of mucosal lesions associated with enteropathogenic and enterohemorrhagic *Escherichia coli* in piglets and the role of plasmid-mediated factors. *Infect Immun* 1989; **57**: 1142-50.
12. Tzipori S, Gunzer F, Donnenberg MS et al. The role of the *eaeA* gene in diarrhea and neurological complications in a gnotobiotic piglet model of enterohemorrhagic *Escherichia coli* infection. *Infect Immun* 1995; **63**: 3621-7.
13. Pai CH, Kelly JK, Meyers GL. Experimental infection of infant rabbits with verotoxin-producing *Escherichia coli*. *Infect Immun* 1986; **51**: 16-23.
14. Yin X, Feng Y, Wheatcroft R et al. Adherence of *Escherichia coli* O157:H7 to epithelial cells in vitro and in pig gut loops is affected by bacterial culture conditions. *Can J Vet Res* 2011; **75**: 81-8.
15. Grys TE, Siegel MB, Lathem WW, Welch RA. The StcE protease contributes to intimate adherence of enterohemorrhagic *Escherichia coli* O157:H7 to host cells. *Infect Immun* 2005; **73**: 1295-303.
16. Kenny B, Abe A, Stein M, Finlay BB. Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. *Infect Immun* 1997; **65**: 2606-12.
17. Rosenshine I, Ruschkowski S, Finlay BB. Expression of attaching/effacing activity by enteropathogenic *Escherichia coli* depends on growth phase, temperature, and protein synthesis upon contact with epithelial cells. *Infect Immun* 1996; **64**: 966-73.
18. Ebel F, Deibel C, Kresse AU et al. Temperature- and medium-dependent secretion of proteins by Shiga toxin-producing *Escherichia coli*. *Infect Immun* 1996; **64**: 4472-9.

19. Beltrametti F, Kresse AU, Guzman CA. Transcriptional regulation of the esp genes of enterohemorrhagic *Escherichia coli*. *J Bacteriol* 1999; **181**: 3409-18.
20. Monteiro R, Ageorges V, Rojas-Lopez M et al. A secretome view of colonisation factors in Shiga toxin-encoding *Escherichia coli* (STEC): from enterohaemorrhagic *E. coli* (EHEC) to related enteropathotypes. *FEMS Microbiol Lett* 2016; **363**.
21. Gerlach RG, Hensel M. Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* 2007; **297**: 401-15.
22. Schwechheimer C, Kuehn MJ. Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nat Rev Microbiol* 2015; **13**: 605-19.
23. Tan S, Tan HT, Chung MC. Membrane proteins and membrane proteomics. *Proteomics* 2008; **8**: 3924-32.
24. Cordwell SJ. Technologies for bacterial surface proteomics. *Curr Opin Microbiol* 2006; **9**: 320-9.
25. Riley LW, Remis RS, Helgerson SD et al. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med* 1983; **308**: 681-5.
26. Abe H, Tatsuno I, Tobe T et al. Bicarbonate ion stimulates the expression of locus of enterocyte effacement-encoded genes in enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 2002; **70**: 3500-9.
27. Abu-Ali GS, Ouellette LM, Henderson ST et al. Differences in adherence and virulence gene expression between two outbreak strains of enterohaemorrhagic *Escherichia coli* O157 : H7. *Microbiology* 2010; **156**: 408-19.
28. House B, Kus JV, Prayitno N et al. Acid-stress-induced changes in enterohaemorrhagic *Escherichia coli* O157 : H7 virulence. *Microbiology* 2009; **155**: 2907-18.
29. Hirakawa H, Kodama T, Takumi-Kobayashi A et al. Secreted indole serves as a signal for expression of type III secretion system translocators in enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* 2009; **155**: 541-50.
30. Nakanishi N, Tashiro K, Kuhara S et al. Regulation of virulence by butyrate sensing in enterohaemorrhagic *Escherichia coli*. *Microbiology* 2009; **155**: 521-30.
31. Anand SK, Griffiths MW. Quorum sensing and expression of virulence in *Escherichia coli* O157:H7. *Int J Food Microbiol* 2003; **85**: 1-9.
32. Girard F, Dziva F, van Diemen P et al. Adherence of enterohemorrhagic *Escherichia coli* O157, O26, and O111 strains to bovine intestinal explants ex vivo. *Appl Environ Microbiol* 2007; **73**: 3084-90.
33. Xicohtencatl-Cortes J, Monteiro-Neto V, Ledesma MA et al. Intestinal adherence associated with type IV pili of enterohemorrhagic *Escherichia coli* O157:H7. *J Clin Invest* 2007; **117**: 3519-29.
34. Taniguchi Y, Choi PJ, Li GW et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010; **329**: 533-8.
35. Chagnot C, Agus A, Renier S et al. In vitro colonization of the muscle extracellular matrix components by *Escherichia coli* O157:H7: the influence of growth medium, temperature and pH on initial adhesion and induction of biofilm formation by collagens I and III. *PLoS One* 2013; **8**: e59386.
36. Wang CH, Koch AL. Constancy of growth on simple and complex media. *J Bacteriol* 1978; **136**: 969-75.
37. Skurnik M, Bolin I, Heikkinen H et al. Virulence plasmid-associated autoagglutination in *Yersinia* spp. *J Bacteriol* 1984; **158**: 1033-6.
38. Michiels T, Vanooteghem JC, Lambert de Rouvroit C et al. Analysis of virC, an operon involved in the secretion of Yop proteins by *Yersinia enterocolitica*. *J Bacteriol* 1991; **173**: 4994-5009.

39. Hoe NP, Goguen JD. Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. *J Bacteriol* 1993; **175**: 7901-9.
40. Falconi M, Colonna B, Prosseda G et al. Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of virF promoter to transcriptional repressor H-NS. *EMBO J* 1998; **17**: 7033-43.
41. Sim M, Koirala S, Picton D et al. Growth rate control of flagellar assembly in *Escherichia coli* strain RP437. *Sci Rep* 2017; **7**: 41189.
42. Shokri A, Sanden AM, Larsson G. Growth rate-dependent changes in *Escherichia coli* membrane structure and protein leakage. *Appl Microbiol Biotechnol* 2002; **58**: 386-92.
43. McHugh JP, Rodriguez-Quinones F, Abdul-Tehrani H et al. Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis. *J Biol Chem* 2003; **278**: 29478-86.
44. Lin XM, Wu LN, Li H et al. Downregulation of Tsx and OmpW and upregulation of OmpX are required for iron homeostasis in *Escherichia coli*. *J Proteome Res* 2008; **7**: 1235-43.

CHAPTER V: OUTER MEMBRANE PROTEIN A AS A POTENCIAL ANTIGEN CANDIDATE ON THE MODULATION OF ENTEROHAEMORRHAGIC *E. COLI* O157:H7 INTESTINAL COLONISATION

Outer membrane protein A as a potencial antigen candidate on the modulation of Enterohaemorrhagic *E. coli* O157:H7 intestinal colonisation

Ricardo MONTEIRO [1,2], Nelly Caccia [1], Valentin Ageorges [1], Dedier Viala [3], Michel HÉBRAUD [1,3], Valérie LIVRELLI [4], Mariagrazia PIZZA [2], Alfredo PEZZICOLI [2,*], Mickaël DESVAUX [1,*]

[1] Université Clermont Auvergne, INRA, UMR454 MEDiS, F-63000 Clermont-Ferrand, France.

[2] GSK, Via Fiorentina 1, 53100, Siena, Italy.

[3] INRA, Plate-Forme d'Exploration du Métabolisme, F-63122 Saint-Genès Champanelle, France.

[4] Centre de Recherche en Nutrition Humaine Auvergne, UMR UCA INSERM U1071, USC-INRA 2018, Clermont Université - Université d'Auvergne, Faculté de Pharmacie, CHU Clermont-Ferrand, Service Bactériologie Mycologie Parasitologie, Clermont-Ferrand, France

*Corresponding authors: Dr Alfredo PEZZICOLI, GSK, Via Fiorentina 1, 53100 Siena, Italy, E-mail: alfredo.x.pezzicoli@gsk.com. Dr Hab Mickaël DESVAUX-LENÔTRE, INRA (Institut National de la Recherche Agronomique) Auvergne-Rhône-Alpes, Centre de Recherche de Clermont-Ferrand, UMR454 MEDiS (Microbiologie, Environnement Digestif, Santé), Site de Theix, F-63122 Saint-Genès Champanelle, France. Tel.: +33 (0)4 73 62 47 23, Fax: +33 (0)4 73 62 45 81, E-mail: mickael.desvaux@inra.fr.

ABSTRACT

Escherichia coli are a versatile pathogen capable of causing intestinal and extraintestinal infections. The plasticity of its genome has let the evolution of these organisms to pathogen strains able to cause disease and syndromes, which result in a huge global human menace. *E. coli* strains are also a major driver of antibiotic resistance, emphasising the urgent need for new treatment and prevention measures. Antigenic and structural heterogeneity among EHEC colonisation factors has complicated vaccine development efforts. This feature associated with the emerging new strains combining virulence factors from different *E. coli* pathotypes indicates that targeting conserved proteins could be more effective. Here we combined proteomics analysis with functional protein characterisation to identify a group of proteins potentially involved in EHEC O157:H7 adhesion to extracellular matrix and to HT-29 intestinal cell line. One of these proteins, OmpA, characterised as higlhy conserved immunogenic antigen, revealed an important role in the adherent phenotype of EHEC O157:H7 as well promoting bacterial aggregation. Moreover antibodies raised against recombinant OmpA showed to effectively reduce the adhesion of EHEC to HT-29 cell model. The present work point OmpA as a potential antigen to be exploited in the development of a vaccine against intestinal pathogenic *E. coli*.

INTRODUCTION

Enterohaemorrhagic *E. coli* (EHEC) are pathogenic Shiga-toxin encoding *E. coli* (STEC) that are able to cause the typical symptoms from watery diarrhea, bloody diarrhea, up to the haemorrhagic colitis (HC), the haemolytic uremic syndrome (HUS), and/or other extraintestinal sequelae in humans [1]. Little progress has been made in reducing the huge number of infections associated with these pathogens and few interventions to reduce the food contamination and the infectious complications of this enteric disease are available [2]. While the vast majority of EHEC infections are sporadic, they can also lead to major outbreaks worldwide [3]. The prevention of *E. coli* infections is of pressing concern from both the public health and economic perspectives [4]. Indeed, the high range of diseases caused by *E. coli*, associated with high costs to healthcare systems makes the absence of a broadly protective vaccine against pathogenic *E. coli* strains a major problem for modern society. The overall problem was exacerbated when, in April 2014, the World Health Organisation published the first global report on antibiotic resistance revealing serious, worldwide threat to public health [5]. However, attempts to develop a broadly protective and safe vaccine against *E. coli* have not been successful so far, being the genetic and antigenic variability of pathogenic *E. coli* species the principal obstacle. This variability coupled with the difficulty prediction of vaccine coverage and protection, has led to denying numerous promising pre-clinical candidates by human trials [6, 7]. Despite the serotype O157:H7, EHEC infections and associated diseases are also related to other 5 STEC serotypes, O26:H11, O45:H2, O103:H11, O11, O121:H19, O145 and O113:H21 [8].

EHEC pathogenesis is not only limited to toxin-mediated effects. The adherence of EHEC on intestinal epithelial cells is believed to be a first step for developing these diseases and colonisation factors include fimbrial and nonfimbrial adhesins and the type III secretion system [9]. The chromosomal pathogenicity LEE genes encoding proteins secreted via the type III secretion system,

including *E. coli* secreted protein A (EspA), EspB, EspD, the receptor for intimin (Tir) and the primary adhesion intimin, is involved in the intimate attachment to epithelial, that so called A/E lesion [10]. However this phynotype was never observed in clinical studies and intimin mutants still bind to host epithelial cell, suggesting the involvement of other adhesion [8, 11, 12]. Several non-fimbrial EHEC adhesions have been reported in adhesion such plasmid-encoded *toxB*, the EHEC factor for adherence *efa1*, and the chromosomally-encoded adhesins, Iha (*Vibrio cholerae* IrgA homolog), Cah (calcium-binding antigen 43 homolog), and outer membrane protein A [12]. Also fimbrial strutures are implicated in host adhesion including long polar fimbriae and F9 (type 1 pilus homolog), two type IV pili such HCP from EHEC O157 and TFP from EHEC non-O157 and the sorbitol-fermenting EHEC O157:H-plasmid-encoded fimbriae [13-15]. This large antigenic and genetic variability of EHEC has been a major obstacle to the development of an effective vaccine and targeting accessory components encoded by the *E. coli* genome may be insufficient. Indeed, currently vaccine strategis focusing individual pathotypes and targeted major virulence determinants revealed unsuccessful, still broader strategies directed against conserved features of all strains may be more effective [16].

In the current study, we investigated whether EHEC O157:H7 prototype strains lacking some outer membrane proteins are able to bind intestinal ECM proteins and adhere to epithelial cells. Also we evaluate the inhibition of EHEC adhesion by antibodies raised against outer membrane proteins. Our data highlighted OmpA as a potential candidate for preventive strategies agains EHEC instetinal colonisation.

MATERIALS AND METHODS

Cell and culture conditions

The non-toxigenic *E. coli* O157:H7 CM454 (isogenic mutant of EHEC O157:H7 EDL933) was used in this study [17-19]. Bacteria were cultured either in BHI (brain-heart infusion, Becton-Dickinson), DMEM (Dulbecco's modified eagle medium, Gibco) or M9. From -80 °C stock culture previously grown in the respective medium, strains were plated on the relevant agar medium and incubated overnight at 37 °C. A preculture was set up from one bacterial colony grown in the respective nutrient broth medium at 37 °C in an orbital shaker at low speed (70 rpm) till stationary phase. For all the experiments, after 1:100 dilution, bacterial cultures were grown in the same conditions until exponential phase (0.5 OD_{600nm}).

HT-29 human colon adenocarcinoma cell line (ATCC) was expanded in 75 cm^2 flasks using DMEM supplemented with 10% foetal bovine serum, 5 mM galactose and 100 µg/ml of penicillin/streptomycin until until ~80% confluency and were used after third passage. For bacterial infection assays, HT-29 cell line were dissociated with 0.05% trypsin-0.02% EDTA and were seeded at a density of 2.0×10^5 cells/cm^2 and grown during 7 days.

Generation of isogenic deletion mutants

Deletion mutants for the genes coding proteins OmpA, OmpC and OmpX were generated. The gene deletion was introduced using the protocol described by Datsenko and Wanner [20]. In short, a kanamycin-resistance cassette flanked by FRT sites was amplified from the pKD4 plasmid using primers with homologous ends complementary to the flanking sequences of the target gene (supplementary data 1). This PCR product was electroporated in strain already harbouring the pKD46 plasmid expressing the λ-red genes to promote homologous recombination. Positive

colonies were selected on kanamycin resistance and correct deletion of the target gene was assessed by PCR and sequencing.

Bacterial adhesion assay to ECM proteins

Preparation of 96-wells polystyrene microtitre plates (Falcon) surface-coated with ECM proteins was based on a previously described protocol [19, 21]. The ECM proteins consisted of collagen I (Millipore), III (Millipore) and IV (Sigma), Muc1 (Sigma), Muc2 (Sigma), laminin-α2 (Millipore), elastin (Sigma), insoluble and soluble fibronectin (Sigma) and MaxGel (Sigma). BSA (bovine serum albumin, Sigma) was used as a control for specific adhesion to ECM proteins. Basically, ECM proteins were solubilised in 0.1 M carbonate coating buffer (pH 9.6) and 250 µl was dispensed at a saturating concentration (50 µg/ml) to the well surface and incubated overnight at 4 °C. The wells were washed with PBS (phosphate buffered saline, Sigma) containing 0.05% (v/v) Tween 20 (PBST, pH 7.3) at room temperature and used for bacterial adhesion assays.

Bacterial ECM adhesion assay follow the method optimised by Chagnot *et al.* [19]. Cloramphenicol was added and mixed gently at a final concentration of 90 µg/ml to bacterial cultures. This step inhibits the growth and adaptation during the time of contact of bacterial cells with ECM proteins in the adhesion assay. Vigorous shaking, vortexing and centrifugation were avoided to preserve cell surface supramolecular structures potentially involved in adhesion. *E. coli* O157:H7 cell suspension was deposited in relevant protein-coated wells of the microtitre plate using wide-bore tips. Control wells were filled with sterile nutrient medium. Microtiter plates were incubated statically at 37 °C for 3 h. After incubation, bacterial suspension was removed by pipetting. Wells were further first washed with TS (tryptone salt) to remove loosely attached cells. Adherent bacteria were fixed with absolute ethanol for 20 min. Wells were then emptied by pipetting and dried for 30 min prior to 20 min staining with an aqueous-solution of crystal violet (0.1% w/v). Wells were washed a second time with TS to remove the excess of unbound crystal violet dye, and dried for 30 min. The bound

dye was solubilised from stained cells using of an aqueous solution of acetic acid (33% v/v) for 1 min under orbital shaking. Contents of each well were transferred to a clean microtiter plate and absorbance was measured at 595 nm using a microtiter plate reader. The readings were normalised by substracting the average absorbance from control wells.

Proteomic identification of bacterial surface proteins binding specifically to ECM

An original proteomic strategy inspired by the work of Dreisbach *et al*. [22] was developed where the cell surface proteins binding specifically to ECM protein were recovered and identified. First, the outer membrane fraction of *E. coli* O157:H7 CM454 was isolated. Bacterial cells were harvested at exponential phase (0.5 OD_{600nm}) and pelleted by centrifugation (5000 xg during 10 min at 4 °C) prior to resuspension in 0.1 M of Tris-HCl (pH=7.3) supplemented with 1 µl/ml of Dnase. Bacterial cells were disrupted using a french press (2.5 kbar) and cellular debris discarded by centrifugation (5000 g during 10 min at 4 °C). To enhance membrane precipitation, supernatant was diluted to a final concentration of 0.1 M of sodium carbonate during 1 h at 4 °C in agitation. Outer membrane (OM) fraction was isolated by ultracentrifugation of 120 000 g during 1 h at 4 °C. Membrane pellet was washed and resupended in 0.1 M of Tris-HCl (pH=7.3) and protein content was quantified following Bradford method [23].

From the identification of bacterial surface proteins binding specifically to collagen I, the OM fraction was added at a 1:10 ratio (m/m) (1 µg of collagen I for 10 µg of proteins in OM fraction) in the wells of microtiter plates coated with collagen I and incubated for 1 h at 37 °C. Wells were thoroughly washed twice with 0.1 M Tris-HCl to remove unbound material. Then, trypsin (Promega) was added and incubated at 37 °C for 1 h. Prior to LC-MS/MS analysis, samples were stacked on SDS-PAGE, excised and reduced in 100 mM ammonium bicarbonate with 45 mM dithiothreitol for 45 min at 50 °C, and alkylated with 100 mM iodoacetamide for 20 min at room temperature in the dark. The gel was washed in 25 mM ammonium bicarbonate and 5% ACN for 30

min and twice in 25 mM ammonium bicarbonate and 50% ACN for 30 min each. Gel bands were dehydrated with 100% acetonitrile and then reswelled in 100 mM ammonium bicarbonate containing 4 mg of trypsin to digest peptides at 37 °C for 5 h. To stop the tryptic digestion and extract the proteins from the gel, 0.1% TFA in 100% ACN was added. Peptide mixtures were analysed by nanoflow liquid chromatography using the Ultimate 3000 RSLC (Dionex, Voisins le Bretonneux, France) with nanocapillary columns (15 cm long 75 µm internal diameter; Acclaim Pep Map RSLC, Dionex). The solvent gradient was increased linearly from 4% to 90% ACN in 0.5% formic acid at a flow rate of 300 nl min^{-1} for 38 min. The elute was electrosprayed inside an LTQ-VELOS mass spectrometer (Thermo Fisher Scientific, Courtaboeuf, France) through a nanoelectrospray ion source. Thermo Proteome Discoverer v1.3 was used for raw data file processing. The *E. coli* O157:H7 strain EDL933 database (T number T00044) was used for protein identification. The following parameters were set for the searches: peptide mass tolerance of 1.5 Da, fragment mass tolerance of 0.5 Da, and a maximum of two missed cleavages allowed. Variable modifications considered were methionine oxidation (M) and carbamidomethylation (C) of cysteine. A protein was considered to be valid when a minimum of two unique peptides originating from one protein showed statistically significant (p<0.05) Mascot scores (http://www.matrixscience.com).

Bacterial adhesion assay to intestinal epithelial cells

E. coli O157:H7 CM454 and generated isogenic mutants (*ΔompA*, *ΔompC*, *ΔompX*) were grown in DMEM and harvest at OD_{600nm} 0.5 (exponential phase). Bacteria were pelleted and resuspended in infection medium comprised of DMEM supplemented with 1% FBS and 5 mM galactose to prepare the inoculum. The inocula were added to HT-29 cultures previous seeded in a 24 well plate in triplicates using multiplicity of infection 10:1 bacteria per cell, and incubated for 1 h at 37°C, 5% CO_2. Subsequently, the inoculum was removed and cells gently rinsed three times with PBS to

remove any non-adherent bacteria. Then wells are added with 1% saponin (Sigma), which lyses the cells without affecting the bacteria. Cells lysates were plated in LB and colony-forming units (CFU) counted. Percentage of adherent bacterial were standardise with CFU counting from inoculum pre-infection.

For adhesion inhibition assay, *E. coli* O157:H7 CM454 and generated isogenic mutants Δ*ompA* and Δ*ompX* grown in DMEM were harvest at OD_{600nm} 0.5 (exponential phase). Bacteria were pelleted and resuspended in PBS. Bacteria were stained using Oregon green 488 (Sigma) following manufacter instructions during 15 min at 37 °C. Bacteria were pelleted, washed in PBS and resuspended in infection medium (DMEM supplemented with 1% FBS and 5 mM galactose). Bacterial cells were then incubated with specific sera against OmpA or OmpC in triplicates with gradient concentration of 5%, 2.5% and 1.75% of serum during 1 h at 37 °C (inoculum). An anti-whole cell *E. coli* serum was used as control. HT-29 cells grown in black 96 well plates were added with inoculum and incubated at 37 °C during 1 h. Subsequently, the inoculum was removed and cells gently rinsed three times with PBS to remove any non-adherent bacteria. Cells were suspended 1% SDS, 0.1 M NaOH solution and transferred to a clean black bottom 96 well plate. Adherent bacteria fluorescence were mesured at 485/535 nm using Tecan Infinite 200PRO.

Cloning, expression and purification of recombinant proteins

Selected protein fragments were cloned into the pET15b+ vector (Novagen) by the polymerase incomplete primer extension (PIPE) method [24]. In brief, sequences coding for each fragment were amplified by PCR from the *E. coli* O157:H7 CM454 genomic DNA, using the primers listed in supplementary data 2. PCRs generated mixtures of incomplete extension products; by primer design, short overlapping sequences were introduced at the ends of these incomplete extension mixtures, which allowed complementary strands to anneal and produce hybrid vector-insert combinations. *Escherichia coli* HK100 cells [25] were then transformed with vector-insert hybrids.

Single ampicillin-resistant colonies were selected and checked for the presence of the recombinant plasmid by PCR. Plasmids from positive clones were isolated and subcloned into competent *E. coli* BL21 (DE3) cells (Novagen). EnPresso® B growth system (Sigma) was used for expression of recombinants proteins. Recombinant proteins were over expressed using EnPresso® B-biosilta growth kit (Sigma) according to the manufacturer's instructions. Briefly, positive clones were cultured o diluted 1:100 in 75 ml of expression medium contained in Corning® Erlenmeyer baffled cell culture flasks (Sigma), and grown overnight at 30 °C, under aeration (160 rpm). Expression medium consisted of 3 medium tablets (EnPresso® B kit – biosilta) dissolved in H_2O, and it was supplemented with amylase (EnPresso® B kit – biosilta) and ampicillin (final concentration 100 µg/ml). On the second day, 1.5 booster tables and amylase (EnPresso® B kit – biosilta) were added to the bacterial culture in order to maintain cell viability, and protein expression was induced by the addition of 1 mM IPTG (Sigma). Bacteria culture was grown at 25 °C, under aeration (160 rpm). Bacteria were pelleted and resuspended in 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Immidazole, pH=8, buffer, supplemented with protease inhibitors (cOmplete™, EDTA-free Protease Inhibitor Cocktail, Roche), bacteria cells were lysed by sonication. Cell lysate was centrifuge at 15 000 xg, for 1 h, at 4°C in order to collect the soluble fraction containing the expressed protein. Recombinant proteins were purified by affinity chromatography. Soluble fraction was filtered using 0.22 µm filter, and then loaded on 5 mL HisTrap FF Crude column (GE Healthcare). After an 20 column volume washing with 50 mM NaH_2PO_4, 300 mM NaCl, 30 mM immidazole, pH=8, buffer supplemented with protease inhibitors, the protein was eluted with 50 mM NaH_2PO_4, 300 mM NaCl, 300 mM immidazole, pH=8, buffer supplemented with protease inhibitors (Roche). For each purification, flow through, washes and elution fractions were separated by SDS-PAGE. Protein content was quantified using the BCA (Themo Fischer Scientific) and purity was checked by SDS-PAGE.

Observation in confocal fluorescent microscopy

E. coli O157:H7 CM454 strains were harvested at OD_{600nm} 0.5, wash and resuspended in PBS. Bacteria were fixed in paraformaldehyde 2% (PFA) for 20 min on a poly-L-lysine-coated slide (Thermo scientific). For infection of HT-29 cells, infected cells were fixed in cold methanol-acetone (50-50%). After a blocking step in 10% normal goat serum (Invitrogen), slides were incubated with anti-OmpA or anti-OmpX mouse serum and then with a goat anti-mouse IgG (Jackson Immuno-Research Laboratories). *E. coli* O157:H7 strain CM454 were localised using rabbit polyclonal antibodies raised against whole cell *E. coli*, and Alexa Fluor488 goat antirabbit IgG (Life Technologies) as secondary antibody. The samples were mounted using the Pro-Long Gold antifade reagent containing the blue-fluorescent nuclear counterstain DAPI (Invitrogen). Images were acquired using a 64 x or 100 x oil objective (1.4 n.a.) mounted on a Zeiss LSM710 confocal microscope. In the pictures the signal from OmpA and OmpX was pseudocoloured in green, while the signals from bacteria are shown in red.

Statistical analysis

Statistical analysis was performed using GraphPad Prism 7. Data of assays result from at least four independent experiments, *i.e.* four biological replicates. On the figures, error bars thus represent the standard deviation from independent experiments. For each experiment, a value was calculated from the average of repetitions performed in triplicate. The mean values from the biological replicates were compared to the mean values obtained with BSA used as a control for ECM adhesion assays (Figure 1); regarding Collagen I coating (Figure 3) and HT-29 cell adhesion (Figure 5), isogenic delection mutants were compared with EHEC CM454; and for sera inhibition assay, different sera concentrations were compared with EHEC CM454 and anti-whole cell *E. coli* (Figure 9). Data were statistically analysed following ANOVA analysis with differences considered

very significant ($p<0.01$, **), highly significant for ($p<0.001$, ***), or very highly significant ($p<0.0001$, ****).

RESULTS

Bacterial culture conditions influence E. coli O157:H7 adhesion to ECM components

In order to evaluate the influence of culture conditions on adherence capability of *E. coli* O157:H7, three different culture media generally used for growth of *E. coli* strains in laboratory were compared. Bacterial adhesion to collagen I, III and IV, mucin 1 and 2, laminin, elastin, insoluble and soluble fibronectin and a complete mixture of ECM molecules MaxGel was accessed. It clearly appeared that adhesion ability greatly depends on the growth media (Figure 1). As for the control wells coated with BSA, *E. coli* O157:H7 grown in the chemically defined medium M9 could not adhere to ECM components. In contrast, bacteria grown in DMEM adhered similarly to the different ECM proteins tested, as well as to BSA, indicating bacterial adhesion besides high, were non-specific. For BHI, high specific bacterial adhesion to both FFC (fibril-forming collagen), i.e. collagens I and III were observed, the adhesion to the remaining ECM tested molecules was low as to BSA control for this culture condition.

Proteomic identification of surface proteins binding specifically to collagen

Following a without-a-priori approach, an original proteomic strategy was developed where the cell surface proteins binding specifically to ECM protein were recovered and identified. Considering *E. coli* O157:H7 showed the highest specific adhesion to collagen I in BHI, the surface proteins responsible for this phenotype were investigated. Collagen bounded proteins were recovered by trypsin hydrolysis. Three conditions were set up, namely, collagen I-only (Figure 2A), OM protein fraction-only (Figure 2C) and collagen incubated with OM protein fraction (Figure 2B). Recovered proteins were separated by SDS-PAGE as show in figure 2. Three specific bands on collagen incubated with OM fraction condition were excised. OmpC, OmpA and OmpX were identified by LC-MS for band 1, 2 and 3 respectivly, suggesting their possible involvment on the specific adhesion to collagen I.

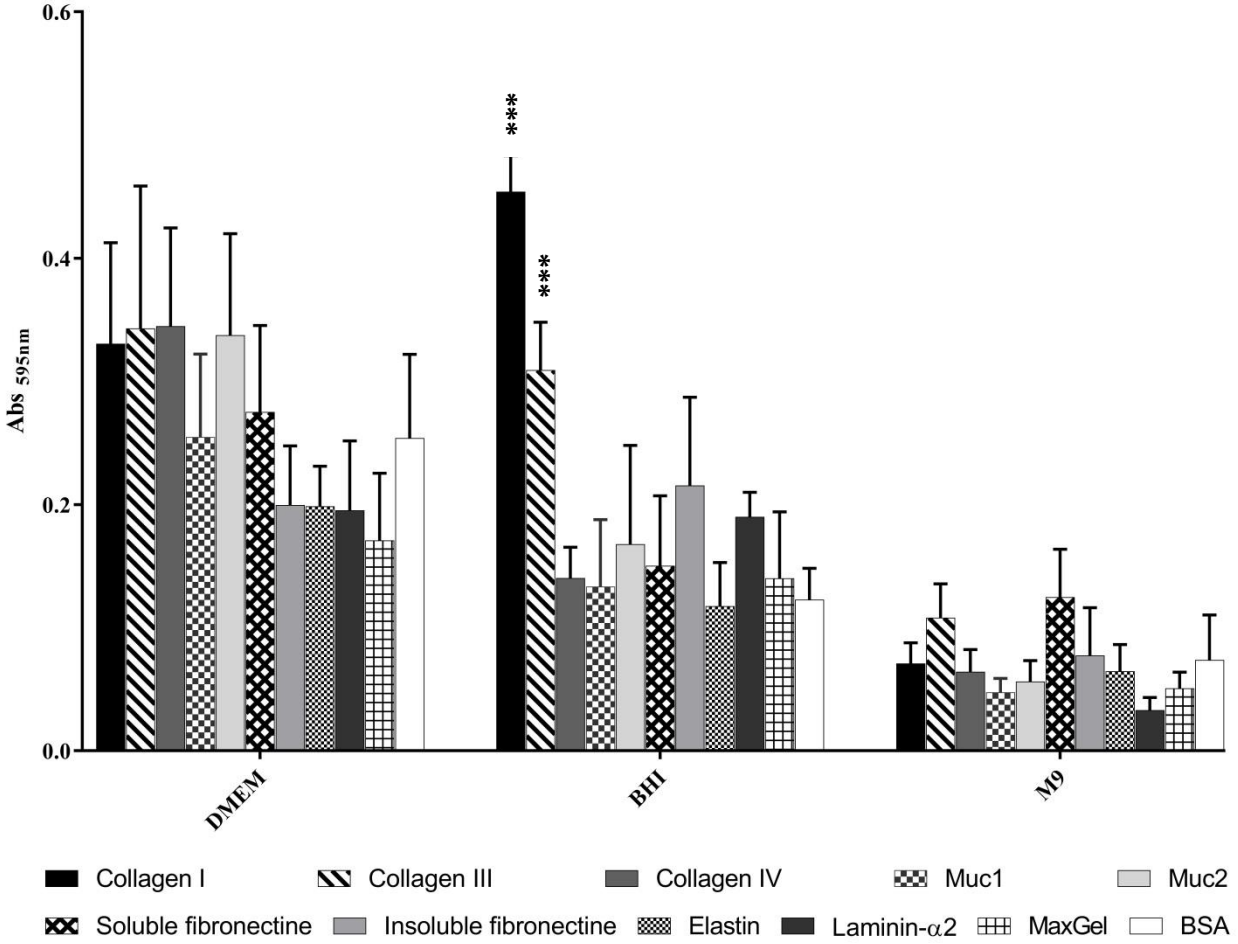
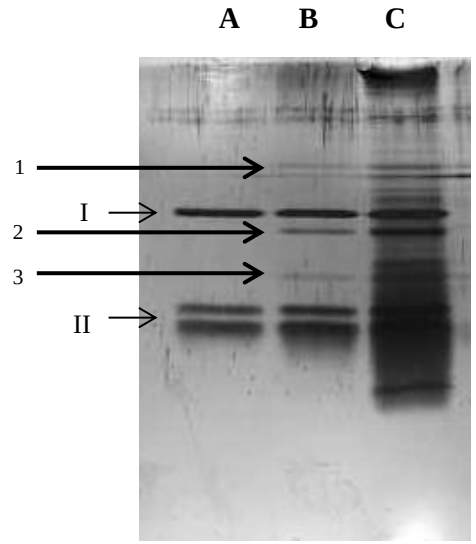


Figure 1: Adhesion to immobilised ECM proteins of *E. coli* O157:H7 strain CM454 grown in DMEM or BHI or M9. Specific bacterial adhesion assay to the main ECM fibrous proteins was performed at 37 °C using BSA as a control and measured by the crystal violet staining method.



Band 1:	Outer membrane protein C		
Accession:	Q8XE41	**Mascot score:**	1340.89
Seq. Coverage [%]:	83.10%	**No. of Peptides**	23
Band 2:	Outer membrane protein A		
Accession:	P0A911	**Mascot score:**	355.31
Seq. Coverage [%]:	29.50%	**No. of Peptides**	5
Band 3:	Outer membrane protein X		
Accession:	P0A919	**Score:**	509.99
Seq. Coverage [%]:	55.00%	**No. of Peptides**	7

Figure 2: SDS-PAGE of the proteins recovered after trypsin digestion in wells with collagen. After coating of OM protein fraction with collagen I, proteins were recovered and separated by SDS-PAGE. Well A corresponds to collagen I-only; well B corresponds to collagen I incubated with OM protein fraction; well C corresponds to OM fraction-only. Bands 1, 2 and 3 are specific of collagen I-OM proteins. Band I and II corresponds to trypsin and autohydrolised trypsin, respectively. Proteins identified in the specific band are described on the picture.

OmpA, OmpC and OmpX are involved in adhesion to collagen I and adhesion to intestinal epithelial cells

In order to demonstrate the involvement of OmpA, OmpC and OmpX in bacterial adhesion to collagen I, functional genetic analysis was carried out and isogenic deletion mutants were generated in the respective genes. Comparing their adhesion to collagen I with *E. coli* O157:H7 CM454 wt, it appeared that all deletion mutants were affected (Figure 3).

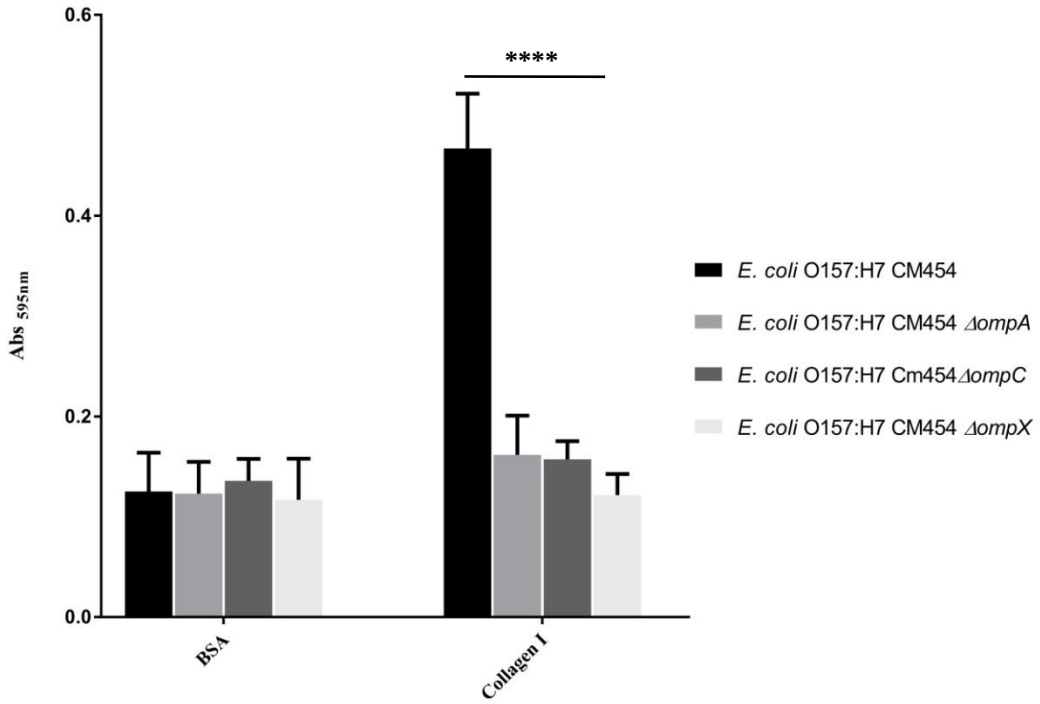


Figure 3: Adhesion to immobilised collagen of *E. coli* O157:H7 strain CM454 and respective delection mutants to *ompA, ompC, ompX* genes. Specific bacterial adhesion to collagen I and III of delection mutants was compared to *E. coli* O157:H7 strain CM454 (black colored bars) and mesured by the crystal violet staining method.

Bacterial adhesion to intestinal epithelial cells was further investigated. HT-29 cells were grown in conditions inducing formation of an apical brush border, with cells differentiated in enterocytes and mucus-producing cells (goblets) [26] (Figure 4A). Seven days post seeding, HT-29 cells presented already a consistent monolayer and around 40% of the cells differentiated into goblet cells producing mucus (green color). At this culture state, the infection behavior of *E. coli* O157:H7 CM454 was evaluated by using confocal microscopy. At one hour incubation with HT-29 cells, *E. coli* O157:H7 CM454 (marked as red color) already adhered to the monolayer, as shown in figure 4B. Bacterial cells tended to adhere better to areas where the mucin staining (marked as green color) was more prominent. Cell adhesion of isogenic deletion mutants were compared with *E. coli* O157:H7 CM454 wild type strain (Figure 5). The deletion of *ompA*, *ompC* or *ompX* genes significantly decreased the adhesion of *E. coli* O157:H7 CM454 to HT-29 cells.

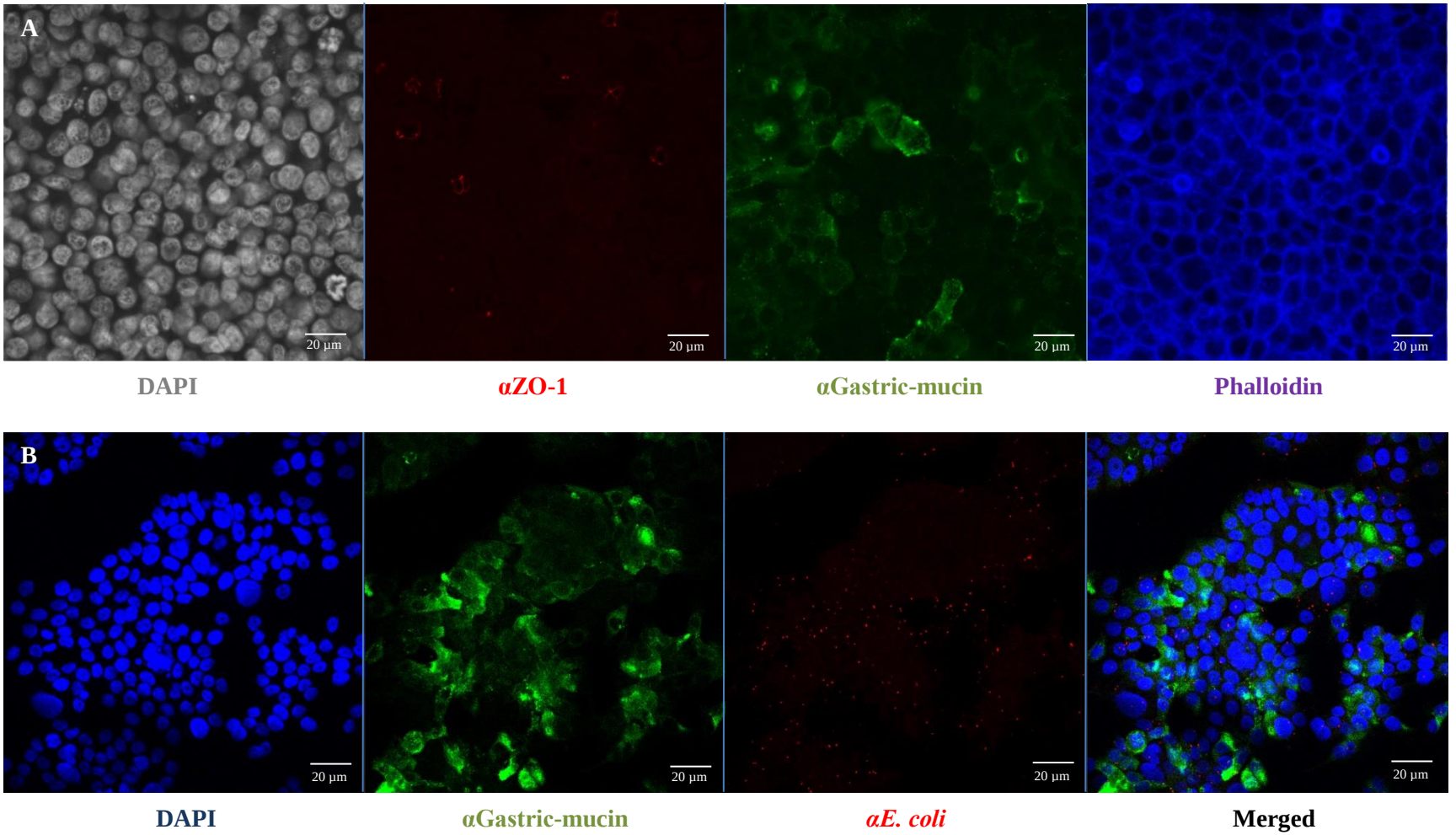


Figure 4: Characterisation of the galactose adapted HT-29 cells monolayer. A) HT-29 cells grown on 24 well plate for 7 days were stained with specific antibodies for Muc1 mucin (αGastric-mucin) and ZO-1. The mucin are stained in green, tight junction in red (ZO-1) and the actin skeleton in blue. DAPI (grey) staining was used to visualise cell nuclei. B) The mucin is stained in green, *E. coli* O157:H7 CM454 in red and cell nuclei in blue.

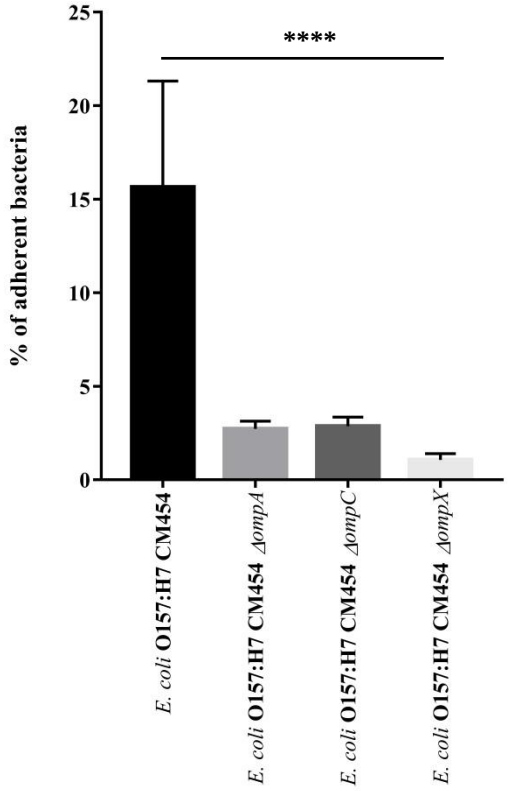
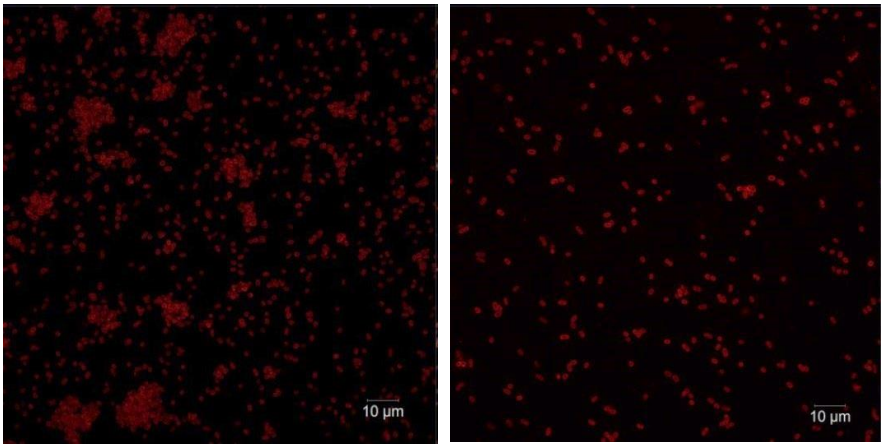


Figure 5: *E. coli* O157: H7 CM454 adhesion assay to HT-29 cells. Adhesion of deletion mutants strains for *ΔompA*, *ΔompC*, *ΔompX* genes was evaluated and compared to the wild type. CFU of adherent bacteria were counted and mutant adhesion percentage was compared with reference *E. coli* O157:H7 CM454 (black colored bar).

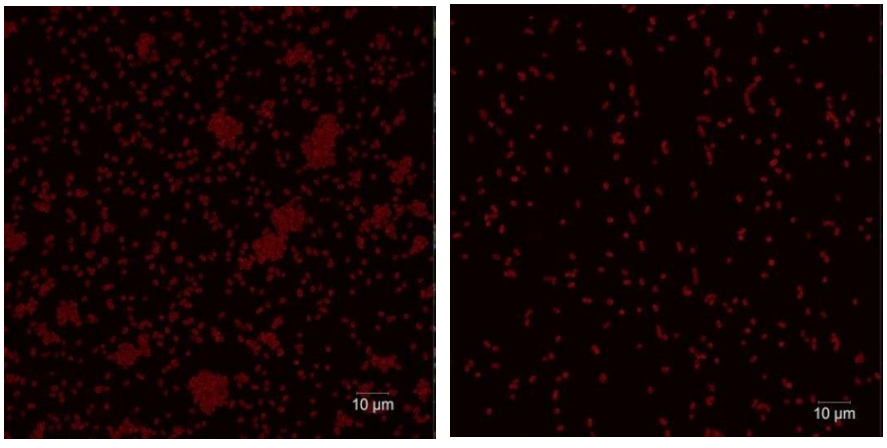
OmpX and OmpA involvement in cell-cell interaction

Despite the fact we were able to express all the proteins in *E. coli* BL21 expression strain, the high number of transmembrane domains of OmpC made impossible to stabilise the recombinant protein in solution. For this reason, antibodies raised against OmpC resulted to be of insufficient quality, leading to unspecific binding to the bacterial surface. On the contrary, antibodies against OmpA and OmpX showed a clear and specific signal by fluorescence microscopy. . Since the deletion of *ompA and ompX* genes correlated with a decreased adhesion to HT29 monolayers, we decided to investigate the localization of the proteins coded by these genes in order to put in relationship the adhesion capability of *E. coli* O157:H7 and their respective localisation during cell-cell interaction.

Fluorescence Microscopy images using anti-OmpA and OmpX antibodies showed that, probably, they perform similar functions in cell-cell interactions and both promote bacterial autoaggregation (Figure 6). Indeed, bacterial aggregates were absent in cultures of *ΔompA* and *ΔompX* strains. It has also been observed that OmpA and OmpX proteins are not equally distributed along the bacterial surface but tend to localize at the interface point between cells. As shown in figure 7A/B, signal from OmpA and OmpX antibodies (green color) are stronger at the cell-cell interface whereas the signal on isolated cells is absent or significantly lower. Interestingly, during the infection of HT-29 cells, the localization of these proteins when bacteria are close to HT-29 cells suggest a specific orientation in direction to the area of cell contact (Figure 8, white arrows).



E. coli* O157:H7 CM454** ***E. coli* O157:H7 CM454 Δ*ompA



E. coli* O157:H7 CM454** ***E. coli* O157:H7 CM454 Δ*ompX

Figure 6: Autoaggregation of *E. coli* O157:H7 CM454. Bacterial cells localised using anti *E. coli* antibodies (red color).

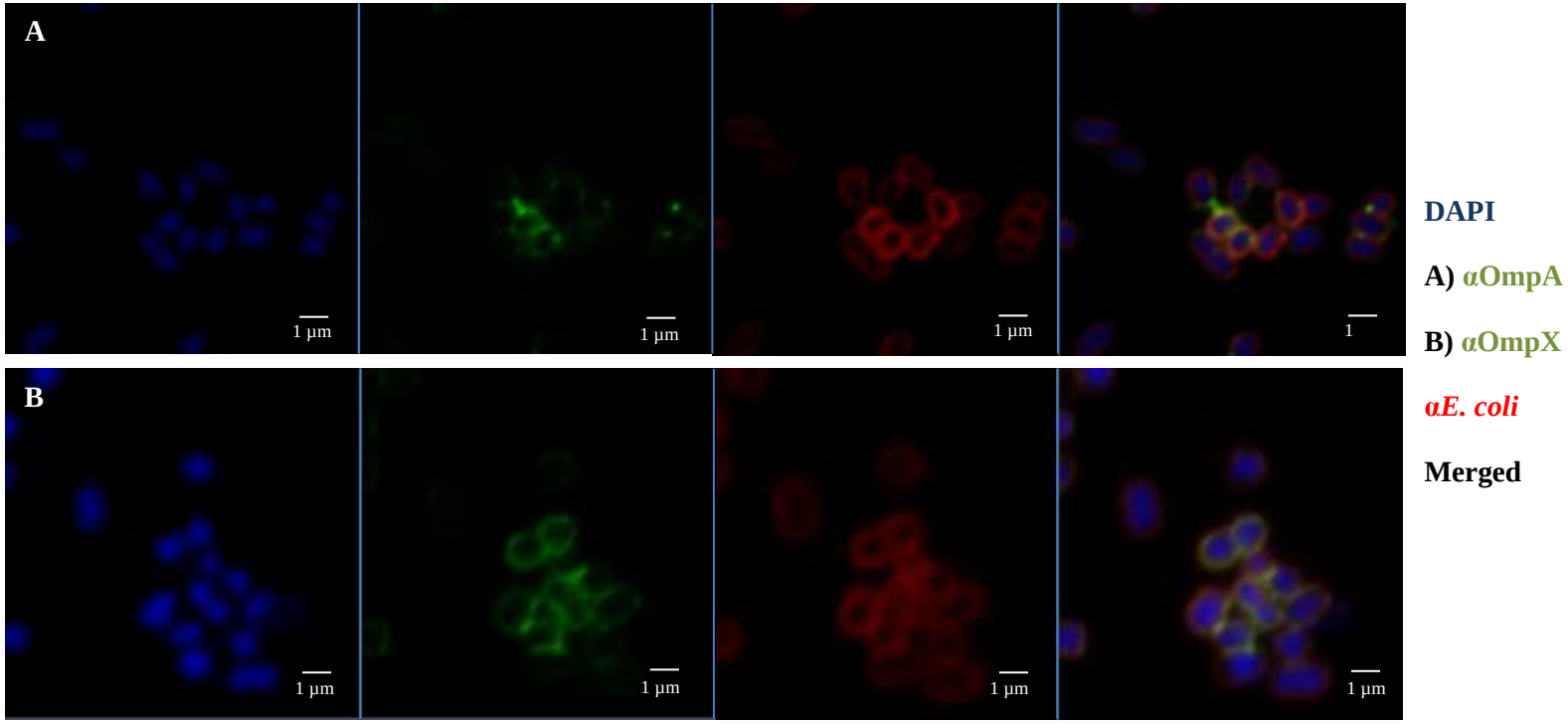


Figure 7: OmpA and OmpX surface localisation on the *E. coli* O157:H7 CM454. Confocal images of (B) OmpA (C) OmpX. Each protein was detected using specific antibodies raised in mouse and visualised using a fluorescent secondary antibody (green). Antibodies to whole-*E.coli* bacteria and a fluorescent secondary antibody (red) and DAPI (blue) were used to visualise bacteria and chromosomal DNA, respectively.

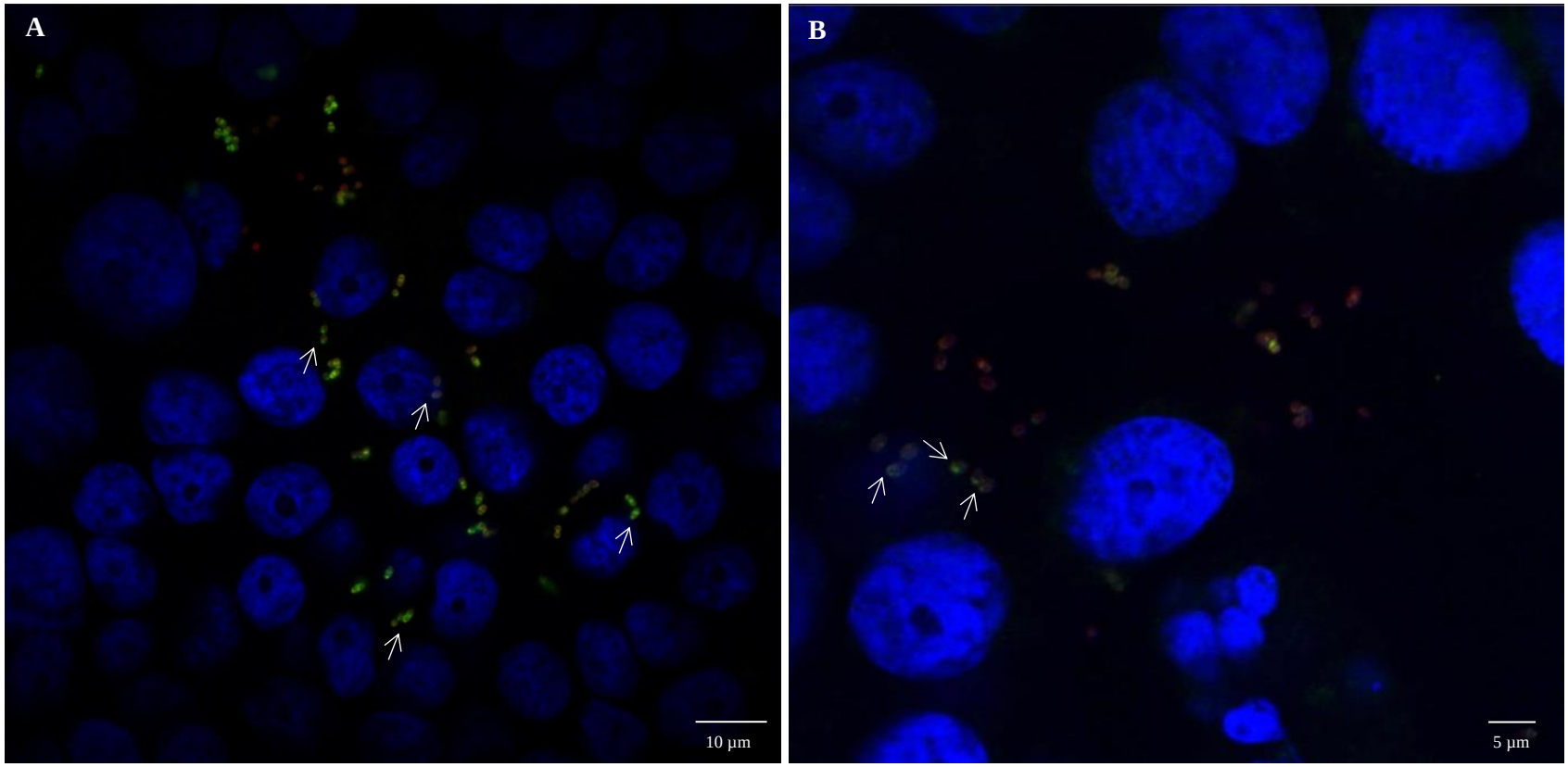


Figure 8: OmpA and OmpX surface localisation during infection of HT-29 cells by *E. coli* O157:H7 CM454. Confocal images of (A) OmpA and (B) OmpX. Each protein was detected using specific antibodies raised in mouse and visualised using a fluorescent secondary antibody (green). Antibodies to whole-*E.coli* bacteria and a fluorescent secondary antibody (red) and DAPI (blue) were used to visualise bacteria and chromosomal DNA, respectively. White arrows point protein orientation to HT-29 cells.

Antibodies anti-OmpA inhibit adhestion to HT-29 cells

Since *E. coli* O157:H7 *ΔompA* and *ΔompX* strains showed lower adhesion ability either to collagen than to the intestinal cell model, the effect of specific antibodies against OmpA and OmpX on bacterial adhesion was evaluated. The efficacy of OmpA antibodies in reducing *E. coli* O157:H7 adhesion to HT-39 intestinal cell model was confirmed by adhesion inhibition assay. Briefly, bacteria were pre-incubated with antibodies against OmpA or OmpX prior to infect HT29 monolayers. Adhesion of *E. coli* O157:H7 bacteria pre-incubated with anti-OmpX antibodies was highly variable and no significant differences could be observed. However, pre-incubation of *E. coli* O157:H7 with antibodies against OmpA significantly prevented its adhesion to HT-29 cells (Figure 9).

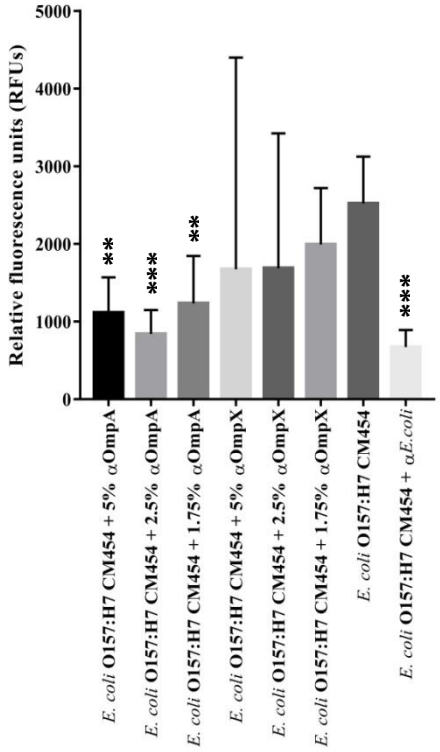


Figure 9: Adhesion inhibition assay. Effect of anti-OmpA and anti-OmpX on *E. coli* O157:H7 CM454 adhesion to HT-29 cells. 5%, 2.5% and 1.75% concentrations of antibodies anti-OmpA or OmpX effect on bacteria-cell adhesion were evaluated and compared with *E. coli* O157:H7 CM454. Antibodies anti-whole *E. coli* cells were used as positive control.

DISCUSSION

Although the adhesion factors implicated in EHEC binding to epithelial cells are under intensive research, the receptors and molecules involved in its recognition and adhesion are poorly characterised or unknown. ECM proteins are generally localised to the epithelial basement membrane, where they are not available for interaction with luminal bacteria. However, interaction with enteric bacterial pathogens can occur during inflammation or the opening of tight junctions [27]. In this study, we assessed the ability of EHEC O157:H7 to bind some major intestinal ECM proteins, such as mucins, collagens, laminin, elastin and fibronectin. Indeed, testing three commonly culture media used to growth EHEC, we have shown different adhesion profiles to ECM proteins. In DMEM non-specific bacterial adhesion to ECM proteins was observed. DMEM was already reported as inducer of bacterial adhesion due to the presence of sodium bicarbonate that induces expression of T3aSS [28]. Some MSCRAMM (microbial surface components recognizing adhesive matrix molecules) proteins can be involved in bacterial adhesion to ECM proteins. In BHI, *E. coli* O157:H7 showed specific adhesion to collagen I and III suggesting the involvement of specific surface molecular factor(s) for adhesion to the main FFCs, such as some outer membrane proteins [9, 29].

In order to identify the potential protein involved in specific adhesion to collagen, an innovative proteomic approach was developed where the bacterial cell-surface proteins binding specifically to ECM protein were recovered and identified. This method proved successful in recovering and identifying some collagen-binding proteins, namely OmpA, OmpC and OmpX. Further characterisation allowed demonstrating OmpA and OmpX were involved in autoaggregation and adhesion to epithelial cells. While OmpA is known to be involved in biofilm formation in *E. coli* [30], OmpX also participates to virulence in *E. coli* [31, 32]. Both were also reported as involved in pathogenesis or adhesion [31-35]. To investigate bacterial adhesion to intestinal epithelial cells, HT-

29 cells were differentiated into enterocytes after the adaptation to galactose. It appeared that *E. coli* O157:H7 could adhere to mucus, especially in areas where it was highly produced. Of note, increasing mucin expression from host inflammatory response was previosuly demonstrated to inhibit *E. coli* O157:H7 adhesion to gut epithelium [36]. OmpA was previously reported as binding to a less complex Hela cell model [33] and it was also suggested to play a role in human meningitis, contributing to invasion of the human brain microvascular endothelial cells [34].

For years the development of strategies to prevent *E. coli* host intestinal colonisation focused in particular strains and their specific virulence factors. Examples include toxin-based vaccines such Shiga-toxin produced by EHEC [37] or heat-labile toxin from ETEC [38]; protein-based vaccines such components and effectors of the T3aSS from EPEC [39, 40] or targeting adhesins or fimbriae from EHEC/ETEC [41-43]. In this study and out of three major OMPs, OmpA appeared as a possible target for preventing intestinal colonisation since antibodies against OmpA could lead to a significant decrease of bacterial adhesion to intestinal epithelial cells. Taken together, OmpA fulfills many prerequisites required for a vaccine candidate since it is soluble and stable and highly prevalent. OmpA was recently proven to be immunogenic since mice were protected from *E. coli* infection after immunisation with recombinant OmpA and cross protected against other pathogens such *Shigella*, *Salmonella*, and *Pseudomonas* [44]. Further investigations in those directions would still be required considering its degree of conservation in numerous pathogenic and commensal Gram-negative bacteria. Especially, the impact of an immunisation on the host microbiome would need to be evaluated [45] as well as the efficacy of different OmpA formulations, e.g. alone or in combination with other broadly prevalent and/or pathotype-specific antigens.

ACKNOWLEDGEMENTS

This work was supported in part by INRA (Institut National de la Recherche Agronomique), GSK and by EU FP7 Marie Curie Actions ITN (Initial Training Network) EID (European Industrial Doctorates) DIS*Co* (A multidisciplinary Doctoral Industrial School on novel preventive strategies against *E. Coli* infections; http://www.discoproject.eu/; n°FP7-PEOPLE-607611) project. Ricardo MONTEIRO is a Marie Curie PhD Research Fellow granted by ITN EID DIS*Co*.

ANNEX

Supplementary data 1

Deletion mutants - primers

OmpA

OmpA-Fw: CTCGTTGGAGATATTCATGGCGTATTTTGGATGATAACGAGGCGCAAAAA
GTGTAGGCTGGAGCTGCTTC

OmpA-Rv: AAAGGCAAAAAAAACCCCGCAGCAGCGGGGTTTTTCTACCAGACGAGAAC
CATATGAATATCCTCCTTAGT

OmpC

Ompc-CmFw: TGCAGTGGCATAAAAAAGCAAATAAAGGCATATAACAGAGGGTTAATAAC
GTGTAGGCTGGAGCTGCTTC

OmpC-CmRv: AAAACAATGAAAAAAGGGCCCGCAGGCCCTTTGTTCGATATCAATCGAGA
CATATGAATATCCTCCTTAGT

OmpX

OmpX-Fw: TTTGATATATTTAAAACTTAGGACTTATTTGAATCACATTTGAGGTGGTT

GTGTAGGCTGGAGCTGCTTC

OmpX-Rv: AAAAACAAAAATCCGCCCCGAGAGGCGGATTTTTTATATCACCAAAGTGA

CATATGAATATCCTCCTTAGT

Supplementary data 2

Overexpression of recombinant proteins

Primers and protein sequences

OmpA

OmpA-Fw: CTGTACTTCCAGGGCGCTCCGAAAGATAACACCTGG

OmpA-Rv: AATTAAGTCGCGTTAAGCTTGCGGCTGAGTTACAAC

APKDNTWYTGAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYVGFEMGYDWLGRMPY
KGSVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTKSNVYGKNHDTGVSPVFAGGV
EYAITPEIATRLEYQWTNNIGDAHTIGTRPDNGMLSLGVSYRFGQGEAAPVVAPAPAPAPEVQTKHF
TLKSDVLFNFNKATLKPEGQAALDQLYSQLSNLDPKDGSVVVLGYTDRIGSDAYNQGLSERRAQSV
VDYLISKGIPADKISARGMGESNPVTGNTCDNVKQRAALIDCLAPDRRVEIEVKGIKDVVTQPQA

OmpX

OmpX-Fw: CTGTACTTCCAGGGCTCCGTAGCTGCGACTTCTAC

OmpX-Rv: AATTAAGTCGCGTTAGAAGCGGTAACCAACACCG

SVAATSTVTGGYAQSDAQGQMNKMGGFNLKYRYEEDNSPLGVIGSFTYTEKSRTASSGDYNKNQY
YGITAGPAYRINDWASIYGVVGVGYGKFQTTEYPTYKHDTSDYGFSYGAGLQFNPMENVALDFSY
EQSRIRSVDVGTWIAGVGYRF

OmpC

OmpC-Fw: CTGTACTTCCAGGGCGGCAACAAATTAGATCTGTACGG

OmpC-Rv: AATTAAGTCGCGTTAGAACTGGTAAACCAGACCCAG

GNKLDLYGKVDGLHYFSDDKSVDGDQTYMRLGFKGETQVTDQLTGYGQWEYQIQGNSAENENNS
WTRVAFAGLKFQDVGSFDYGRNYGVVYDVTSWTDVLPEFGGDTYGSDNFMQQRGNGFATYRNT
DFFGLVDGLNFAVQYQGKNGSVSGEGMTNNGREALRQNGDGVGGSITYDYEGFGIGAAVSSSKRT
DDQNSPLYIGNGDRAETYTGGLKYDANNIYLAAQYTQTYNATRVGSLGWANKAQNFEAVAQYQF
DFGLRPSLAYLQSKGKNLGVINGRNYDDEDILKYVDVGATYYFNKNMSTYVDYKINLLDDNQFTR
DAGINTDNIVALGLVYQF

REFERENCES

1. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 2005; **365**: 1073-86.
2. Walker CL, Black RE. Diarrhoea morbidity and mortality in older children, adolescents, and adults. *Epidemiology and infection* 2010; **138**: 1215-26.
3. Karch H, Bielaszewska M, Bitzan M, Schmidt H. Epidemiology and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Diagnostic microbiology and infectious disease* 1999; **34**: 229-43.
4. Smith JL, Fratamico PM, Gunther NW. Extraintestinal pathogenic *Escherichia coli*. *Foodborne Pathog Dis* 2007; **4**: 134-63.
5. WHO. Antimicrobial resistance: global report on surveillance. *World Health Organization* 2014; **http://www.who.int/drugresistance/documents/surveillancereport/en/**.
6. Brumbaugh AR, Mobley HL. Preventing urinary tract infection: progress toward an effective *Escherichia coli* vaccine. *Expert Rev Vaccines* 2012; **11**: 663-76.
7. Ahmed T, Bhuiyan TR, Zaman K et al. Vaccines for preventing enterotoxigenic *Escherichia coli* (ETEC) diarrhoea. *Cochrane Database Syst Rev* 2013: CD009029.
8. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 1998; **11**: 142-201.
9. Farfan MJ, Torres AG. Molecular mechanisms that mediate colonization of Shiga toxin-producing *Escherichia coli* strains. *Infect Immun* 2012; **80**: 903-13.
10. Donnenberg MS, Tzipori S, McKee ML et al. The role of the eae gene of enterohemorrhagic *Escherichia coli* in intimate attachment in vitro and in a porcine model. *J Clin Invest* 1993; **92**: 1418-24.
11. Dytoc MT, Ismaili A, Philpott DJ et al. Distinct binding properties of eaeA-negative verocytotoxin-producing *Escherichia coli* of serotype O113:H21. *Infect Immun* 1994; **62**: 3494-505.
12. Torres AG, Zhou X, Kaper JB. Adherence of diarrheagenic Escherichia coli strains to epithelial cells. *Infect Immun* 2005; **73**: 18-29.
13. Brunder W, Khan AS, Hacker J, Karch H. Novel type of fimbriae encoded by the large plasmid of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H(-). *Infect Immun* 2001; **69**: 4447-57.
14. Low AS, Holden N, Rosser T et al. Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. *Environ Microbiol* 2006; **8**: 1033-47.
15. Low AS, Dziva F, Torres AG et al. Cloning, expression, and characterization of fimbrial operon F9 from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 2006; **74**: 2233-44.
16. Moriel DG, Rosini R, Seib KL et al. *Escherichia coli*: great diversity around a common core. *MBio* 2012; **3**.
17. Gobert AP, Vareille M, Glasser AL et al. Shiga toxin produced by enterohemorrhagic *Escherichia coli* inhibits PI3K/NF-kappaB signaling pathway in globotriaosylceramide-3-negative human intestinal epithelial cells. *J Immunol* 2007; **178**: 8168-74.

18. Perna NT, Plunkett G, 3rd, Burland V et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 2001; **409**: 529-33.
19. Chagnot C, Agus A, Renier S et al. In vitro colonization of the muscle extracellular matrix components by *Escherichia coli* O157:H7: the influence of growth medium, temperature and pH on initial adhesion and induction of biofilm formation by collagens I and III. *PLoS One* 2013; **8**: e59386.
20. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 2000; **97**: 6640-5.
21. Hennequin C, Forestier C. Influence of capsule and extended-spectrum beta-lactamases encoding plasmids upon *Klebsiella pneumoniae* adhesion. *Res Microbiol* 2007; **158**: 339-47.
22. Dreisbach A, van der Kooi-Pol MM, Otto A et al. Surface shaving as a versatile tool to profile global interactions between human serum proteins and the *Staphylococcus aureus* cell surface. *Proteomics* 2011; **11**: 2921-30.
23. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976; **72**: 248-54.
24. Klock HE, Lesley SA. The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods Mol Biol* 2009; **498**: 91-103.
25. Klock HE, White A, Koesema E, Lesley SA. Methods and results for semi-automated cloning using integrated robotics. *J Struct Funct Genomics* 2005; **6**: 89-94.
26. Phillips TE, Huet C, Bilbo PR et al. Human intestinal goblet cells in monolayer culture: characterization of a mucus-secreting subclone derived from the HT29 colon adenocarcinoma cell line. *Gastroenterology* 1988; **94**: 1390-403.
27. Walia B, Castaneda FE, Wang L et al. Polarized fibronectin secretion induced by adenosine regulates bacterial-epithelial interaction in human intestinal epithelial cells. *Biochem J* 2004; **382**: 589-96.
28. Abe H, Tatsuno I, Tobe T et al. Bicarbonate ion stimulates the expression of locus of enterocyte effacement-encoded genes in enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 2002; **70**: 3500-9.
29. !!! INVALID CITATION !!! {}.
30. Barrios AF, Zuo R, Ren D, Wood TK. Hha, YbaJ, and OmpA regulate *Escherichia coli* K12 biofilm formation and conjugation plasmids abolish motility. *Biotechnol Bioeng* 2006; **93**: 188-200.
31. Vogt J, Schulz GE. The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Structure* 1999; **7**: 1301-9.
32. Meng X, Liu X, Zhang L et al. Virulence characteristics of extraintestinal pathogenic *Escherichia coli* deletion of gene encoding the outer membrane protein X. *J Vet Med Sci* 2016; **78**: 1261-7.
33. Torres AG, Jeter C, Langley W, Matthysse AG. Differential binding of *Escherichia coli* O157:H7 to alfalfa, human epithelial cells, and plastic is mediated by a variety of surface structures. *Appl Environ Microbiol* 2005; **71**: 8008-15.
34. Shin S, Lu G, Cai M, Kim KS. *Escherichia coli* outer membrane protein A adheres to human brain microvascular endothelial cells. *Biochem Biophys Res Commun* 2005; **330**: 1199-204.

35. Wang H, Li Q, Fang Y et al. Biochemical and functional characterization of the periplasmic domain of the outer membrane protein A from enterohemorrhagic *Escherichia coli*. *Microbiol Res* 2016; **182**: 109-15.
36. Xue Y, Zhang H, Wang H et al. Host inflammatory response inhibits *Escherichia coli* O157:H7 adhesion to gut epithelium through augmentation of mucin expression. *Infect Immun* 2014; **82**: 1921-30.
37. Bitzan M, Poole R, Mehran M et al. Safety and pharmacokinetics of chimeric anti-Shiga toxin 1 and anti-Shiga toxin 2 monoclonal antibodies in healthy volunteers. *Antimicrob Agents Chemother* 2009; **53**: 3081-7.
38. Glenn GM, Villar CP, Flyer DC et al. Safety and immunogenicity of an enterotoxigenic *Escherichia coli* vaccine patch containing heat-labile toxin: use of skin pretreatment to disrupt the stratum corneum. *Infect Immun* 2007; **75**: 2163-70.
39. Gansheroff LJ, Wachtel MR, O'Brien AD. Decreased adherence of enterohemorrhagic *Escherichia coli* to HEp-2 cells in the presence of antibodies that recognize the C-terminal region of intimin. *Infect Immun* 1999; **67**: 6409-17.
40. Horne C, Vallance BA, Deng W, Finlay BB. Current progress in enteropathogenic and enterohemorrhagic *Escherichia coli* vaccines. *Expert Rev Vaccines* 2002; **1**: 483-93.
41. Roy K, Hamilton D, Ostmann MM, Fleckenstein JM. Vaccination with EtpA glycoprotein or flagellin protects against colonization with enterotoxigenic *Escherichia coli* in a murine model. *Vaccine* 2009; **27**: 4601-8.
42. Harro C, Sack D, Bourgeois AL et al. A combination vaccine consisting of three live attenuated enterotoxigenic *Escherichia coli* strains expressing a range of colonization factors and heat-labile toxin subunit B is well tolerated and immunogenic in a placebo-controlled double-blind phase I trial in healthy adults. *Clin Vaccine Immunol* 2011; **18**: 2118-27.
43. Riquelme-Neira R, Rivera A, Saez D et al. Vaccination with DNA Encoding Truncated Enterohemorrhagic *Escherichia coli* (EHEC) Factor for Adherence-1 Gene (efa-1') Confers Protective Immunity to Mice Infected with *E. coli* O157:H7. *Front Cell Infect Microbiol* 2015; **5**: 104.
44. Guan Q, Wang X, Wang X et al. Recombinant outer membrane protein A induces a protective immune response against *Escherichia coli* infection in mice. *Appl Microbiol Biotechnol* 2015; **99**: 5451-60.
45. Hays MP, Ericsson AC, Yang Y, Hardwidge PR. Vaccinating with conserved *Escherichia coli* antigens does not alter the mouse intestinal microbiome. *BMC Res Notes* 2016; **9**: 401.

CHAPTER VI: CONCLUDING REMARKS

CONCLUDING REMARKS

Pathogenic *E. coli* strains represent a major health issue associated with social and financial implications. From a medical point of view and besides diarrheal illness, they are able to cause a wide range of diseases, such as UTIs, sepsis, neonatal meningitis, intra-abdominal infections, nosocomial pneumonia, osteomyelitis, cellulitis and wound infections. Considering the increasing antimicrobial resistance, the number of recurrent infections and the high incidence in susceptible subpopulations, the prevention of *E. coli*-induced diseases is of high clinical importance and can have some significant economical impact [1]. Especially, diarrheal illnesses are a severe public health problem and a major cause of morbidity and mortality in infants and young children [2]. African, Asian and Latin American low- and middle-income countries are the most affected with diarrheal diseases occurring more often with lethal outcomes mainly due to poor living conditions (inadequate water supplies, poor environmental hygiene and sanitation, and insufficient education) [3].

E. coli strains involved in diarrheal diseases, particularly EHEC and ETEC strains, are one of the most important of the various etiological agents of diarrhea [4]. Strains have evolved through by the acquisition of genes, through horizontal gene transfer, which have successfully persisted in the host [1, 3, 4]. The 2011 *E. coli* O104:H4 German/France outbreak updated the understanding of evolution and genome plasticity in intestinal pathogenic *E. coli*. Combining key virulence factors from EAEC and EHEC, this strain emerged as a highly virulent lineage [5]. This outbreak proved that targeting components encoded by individual pathotypes of *E. coli* may be insufficient to prevent the emergence of new pathogenic lineages, and broader strategies directed against conserved features of all strains may be more effective [6]. The efforts to prevent diarreagenic diseases produced by intestinal pathogenic *E. coli* should be strength, not only for the EHEC infections but to the other pathotypes that are becoming relevant agents in the public health. Importantly, treatments and preventive intervetion should also target non-O157 EHECs serotypes, ETEC, and other pathotypes as DAEC, EIEC, and EAEC for which vaccines development and discovery studies are scarce or nulls [7].

Vaccination is the most effective medical intervention ever introduced to prevent infectious diseases. From using inactivated microorganisms to modern recombinant antigens, vaccine efficacy and safety have continuously improved and helped to save millions of lives [8]. Reverse vaccinology has demonstrated to be a powerful tool where conventional approaches to vaccine development had failed, demonstrating what genomic era applied to microbiology could bridge with

the medical field [9-11]. From the very first application of the reverse vaccinology, genomics, transcriptomics and proteomics have integrated the vaccine research in a very tight synergy, boosting each other over the years. Despite the large information given by the genome, the characterization of recombinant protein subunits and glycoconjugation through the analysis of their sequences, post-translational modifications and conjugation site identifications are only possible using proteomics methods [12, 13]. Proteomics largy contribute to supplement genomic data by helping in the interpretation of genomic sequences, such as the identification of protein regions that are absent from or incorrectly represented in current gene annotations or wrongly annotated in term of translation initiation sites from annotated ORFs [13].

In the present research project, differential proteomic high-throughput analysis integrated together with bioinformatic studies and protein functional characterisation allowed to identify OmpA as important adhesion potentially involved EHEC infection and intestinal colonisation.

Previous studies have postulated that intestinal pathogenic *E. coli* have adapted to the human body by developing a sophisticated network of virulence and colonisation factors working in a concert manner [14-17]. Indeed the central fact that the EHEC A/E lesions have never been reported from clinical studies [4] highlights the necessity to better understand the conditions in which experimental studies are being performed. The EHEC behavior in different culture conditions still unclear and only few studies have addressed this environmental question [14, 18-22].

Our data suggest that *E. coli* O157:H7 is adapting its set of secreted proteins in order to face different environmental conditions. For instance EHEC expresses its full secretion potential when grown in minimal medium or induces virulence factors when grown in nutrient rich conditions such DMEM or even oversecrets proteins into extracellular medium or into outer membrane vesicles. Nonetheless deciphering the extracytoplasmic protein expression landscape revealed tricky, especially surface exposed proteins. In chapter III, we demonstrated that biotinylation is an important method to explore the proteosurfaceome. Indeed our findings made evident that different biotinylation reagents could be used with distinct purposes. For instance, reagents of different sizes appeared useful to explore more or less deeply the bacterial cell envelope and label proteins on the external side of the OM, the internal side of the OM, periplasm and/or periplasmic side of IM. The development of a more diverse range of biotinylation reagents commercially available is necessary; especially some reagents truly impermeable to the bacterial OM to investigate the protein exposed the outer most cell envelope layer.

These proteomic achievements in an *E. coli* model strain revealed crucial for the objectives established in the chaper IV. Our analysis of EHEC secretome in different culture conditions

demonstrated the secretion adaptation to surrounding conditions. Indeed EHEC dramatically modulate the level of protein expression in the secretome when grown in M9, DMEM or BHI. We highlighted, particularly, the differential expression of T3SS, an important mechanism of virulence and colonisation in EHEC. With this chapter we evidenced the complexity of EHEC expression and we propose further investigations using bacterial culture conditions as close possible to the real intestinal conditions. The influence on the secretome of other factors relevant to the InPEC infection and lifestyle outside the host would need to be further considered, such as different temperature, different pH, different growth rate and/or aerobic versus microanaerobic/anaerobic conditions.

Also receptors and mechanisms that trigger EHEC infection at the intestinal tract remain unclear [23]. At this point, we hypothesized that extracellular matrix molecules could act as receptors for the initial attachment of EHEC to the intestinal ephitelium. The ECM proteins comprise a complex protein structure involved in several biological proceses, particularly act as a barrier and support for epithelial cells and that are responsible for the development, growth, and maintenance of tissues [24]. The composition of ECM differs from organ to organ, but at the intestinal ephitelium consists mainly in fibronectin, collagen types I, to XV, and laminin [25]. ECM proteins have been shown to act as a substrate for bacterial adherence to eukaryotic cells are commonly recognized by bacterial adhesins [25-27], but in a living organisms how can EHEC have contact with intestinal ECM? Indeed, ECM proteins generally are localized to the basement membrane, but during an event of inflammation or the opening of tight junctions, they can be exposed to enteric bacterial pathogens [28]. Hence, binding to ECM proteins may facilitate colonisation, invasion, and/or signaling by intestinal pathogens. When bacteria were grown in rich undefined condition such BHI, specific binding to collagen I and III culture medium could be demonstrated as involving OmpA. Also, an innovative proteomic strategy to identify bacterial surface proteins specifically involved in ECM protein binding was developed. *E. coli* O157:H7 isogenic mutants lacking OmpA, OmpC or OmpX, undeniably have lowered the adhesion to ECM, indicating that these proteins are involved in ECM proteins adhesion. Intestinal epithelium cell assay demonstrated OmpA and OmpX were involved in adhesion and antibodies against OmpA allowed inhibiting it.

Identifying broadly conserved and protective antigens independent of strain-specific colonisation factors could be, then crucial on preventing intestinal pathogenic *E. coli*. OmpA is highly conserved not only in diverse InPEC isolates, but also in commensal *Proteobacteria*, other *E. coli* and *Shigella* strains, sharing ~99 % identity [6, 29]. Considering the sequence similarities of OmpA from *Salmonella* and *Pseudomonas* could make this protein a possible antigen among these pathogenic

species [30]. Nonetheless, further investigations will be required to confirm the potent role of OmpA in intestinal colonisation, using animal models. Respective to vaccine development, a full set of immunization experiments are still required as well understanding the intestinal homeostasis, especially on the commensal microbiota considering OmpA conservation in numerous diderm-LPS bacteria with a high level of protein similarity. In a recent study, mouse immunized with highy conserved MipA, Skp, or ETEC_2479 proteins revealed no significant changes on mouse health, behavior, or rate of weight gain following intranasal vaccination [31]. Although no significant effect was observed in mouse model, it is known that alteration of commensal microbiota can influence susceptibility to gastrointestinal disease [32] and vaccine efficacy [33] in humans. This further stress the need for complementary and comprehensive studies on the human microbiome. The results presented in this project, showed the significance of quantitative proteomics on antigen discovery and vaccine development in synergy with genomic approaches such reverse vaccinology. Here we suggest that OmpA fulfills many of the requirements for a vaccine candidate and it can fit a future vaccine formulation in combination with other conserved proteins and/or pathotype-specific antigens.

REFERENCES

1. Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli.* Nat Rev Microbiol, 2004. **2**(2): p. 123-40.
2. *World Health Organization, WHO Press; Geneva, Switzerland: 2012. World Health Statistics.*
3. Croxen, M.A., et al., *Recent advances in understanding enteric pathogenic Escherichia coli.* Clin Microbiol Rev, 2013. **26**(4): p. 822-80.
4. Nataro, J.P. and J.B. Kaper, *Diarrheagenic Escherichia coli.* Clin Microbiol Rev, 1998. **11**(1): p. 142-201.
5. Karch, H., et al., *The enemy within us: lessons from the 2011 European Escherichia coli O104:H4 outbreak.* EMBO Mol Med, 2012. **4**(9): p. 841-8.
6. Moriel, D.G., et al., *Escherichia coli: great diversity around a common core.* MBio, 2012. **3**(3).
7. Bouzari, S., et al., *Immune response against adhesins of enteroaggregative Escherichia coli immunized by three different vaccination strategies (DNA/DNA, Protein/Protein, and DNA/Protein) in mice.* Comp Immunol Microbiol Infect Dis, 2010. **33**(3): p. 215-25.
8. van Panhuis, W.G., et al., *Contagious diseases in the United States from 1888 to the present.* N Engl J Med, 2013. **369**(22): p. 2152-8.
9. Rappuoli, R., *Reverse vaccinology, a genome-based approach to vaccine development.* Vaccine, 2001. **19**(17-19): p. 2688-91.
10. Rappuoli, R., et al., *Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design.* J Exp Med, 2016. **213**(4): p. 469-81.
11. Rappuoli, R., *Reverse vaccinology.* Curr Opin Microbiol, 2000. **3**(5): p. 445-50.
12. Donnarumma, D., et al., *The role of structural proteomics in vaccine development: recent advances and future prospects.* Expert Rev Proteomics, 2016. **13**(1): p. 55-68.
13. Biagini, M., F. Bagnoli, and N. Norais, *Surface and Exoproteomes of Gram-Positive Pathogens for Vaccine Discovery.* Curr Top Microbiol Immunol, 2017. **404**: p. 309-337.
14. Abe, H., et al., *Bicarbonate ion stimulates the expression of locus of enterocyte effacement-encoded genes in enterohemorrhagic Escherichia coli O157:H7.* Infect Immun, 2002. **70**(7): p. 3500-9.
15. Barnett Foster, D., *Modulation of the enterohemorrhagic E. coli virulence program through the human gastrointestinal tract.* Virulence, 2013. **4**(4): p. 315-23.
16. Farfan, M.J. and A.G. Torres, *Molecular mechanisms that mediate colonization of Shiga toxin-producing Escherichia coli strains.* Infect Immun, 2012. **80**(3): p. 903-13.
17. Nguyen, Y. and V. Sperandio, *Enterohemorrhagic E. coli (EHEC) pathogenesis.* Front Cell Infect Microbiol, 2012. **2**: p. 90.
18. Kenny, B., et al., *Enteropathogenic E. coli (EPEC) transfers its receptor for intimate adherence into mammalian cells.* Cell, 1997. **91**(4): p. 511-20.
19. Ebel, F., et al., *Temperature- and medium-dependent secretion of proteins by Shiga toxin-producing Escherichia coli.* Infect Immun, 1996. **64**(11): p. 4472-9.
20. Rosenshine, I., S. Ruschkowski, and B.B. Finlay, *Expression of attaching/effacing activity by enteropathogenic Escherichia coli depends on growth phase, temperature, and*

protein synthesis upon contact with epithelial cells. Infect Immun, 1996. **64**(3): p. 966-73.
21. Chagnot, C., et al., *In vitro colonization of the muscle extracellular matrix components by Escherichia coli O157:H7: the influence of growth medium, temperature and pH on initial adhesion and induction of biofilm formation by collagens I and III.* PLoS One, 2013. **8**(3): p. e59386.
22. House, B., et al., *Acid-stress-induced changes in enterohaemorrhagic Escherichia coli O157 : H7 virulence.* Microbiology, 2009. **155**(Pt 9): p. 2907-18.
23. Nataro, J.P., et al., *Diarrheagenic Escherichia coli infection in Baltimore, Maryland, and New Haven, Connecticut.* Clin Infect Dis, 2006. **43**(4): p. 402-7.
24. Berrier, A.L. and K.M. Yamada, *Cell-matrix adhesion.* J Cell Physiol, 2007. **213**(3): p. 565-73.
25. Dorsey, C.W., et al., *Salmonella enterica serotype Typhimurium MisL is an intestinal colonization factor that binds fibronectin.* Mol Microbiol, 2005. **57**(1): p. 196-211.
26. Farfan, M.J., et al., *Long polar fimbriae of enterohemorrhagic Escherichia coli O157:H7 bind to extracellular matrix proteins.* Infect Immun, 2011. **79**(9): p. 3744-50.
27. Farfan, M.J., K.G. Inman, and J.P. Nataro, *The major pilin subunit of the AAF/II fimbriae from enteroaggregative Escherichia coli mediates binding to extracellular matrix proteins.* Infect Immun, 2008. **76**(10): p. 4378-84.
28. Walia, B., et al., *Polarized fibronectin secretion induced by adenosine regulates bacterial-epithelial interaction in human intestinal epithelial cells.* Biochem J, 2004. **382**(Pt 2): p. 589-96.
29. Moriel, D.G., et al., *A Novel Protective Vaccine Antigen from the Core Escherichia coli Genome.* mSphere, 2016. **1**(6).
30. Li, C., et al., *Identification of a novel vaccine candidate by immunogenic screening of Vibrio parahaemolyticus outer membrane proteins.* Vaccine, 2014. **32**(46): p. 6115-21.
31. Hays, M.P., et al., *Vaccinating with conserved Escherichia coli antigens does not alter the mouse intestinal microbiome.* BMC Res Notes, 2016. **9**(1): p. 401.
32. Zhang, Q., et al., *Commensal bacteria direct selective cargo sorting to promote symbiosis.* Nat Immunol, 2015. **16**(9): p. 918-26.
33. Nakaya, H.I. and O. Bruna-Romero, *Is the gut microbiome key to modulating vaccine efficacy?* Expert Rev Vaccines, 2015. **14**(6): p. 777-9.

ACKNOWLEDGEMENTS

First, I would like to thanks the European Commission for the funding support through the Marie Curie Actions and also to supporting young researchers in the begin of their careers.

I thank also to the two main institutions involved in the project. In INRA, to Dr. Régine Talon as Unit director for the intire availabily of unit resources in the development of the project. In GSK, to Dr. Marco Soriani and later Dr. Mariagrazzia Pizza for coordinating the project.

I thank to the University of Rome (UNIROMA3), to Prof. Paolo Visca for helping and guiding me through the Italian academic process.

I would like to express my sincere gratitude to my advisor Dr. Mickaël Desvaux for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. Particularly I really thank the friendship we have built during these last years.

I would like also to thank my co-superviser in GSK, Dr. Alfredo Pezzicoli for provided me an opportunity to join his lab as PhD, and who gave access to the laboratory and research facilities. Without his precious support it would not be possible to conduct this research in GSK.

I thank my fellow labmates in France and Italy: Nelly, Geoffrey, Ingrid, Tiago, Carine, Sylvie, Jean Paul, Alexandra, Alexandre, Valentin, Audrey, Carlo, Martina, Zaira, Vanessa, Elena and Pietro for the stimulating discussions and for all the fun we have had in the last three years.

Last but not the least, I would like to thank my family, especially to my grandfather Carlos and Angela for supporting me, even though the distance and all time spent abroad.

AUTHOR DISCLOSURE

Ricardo Monteiro, PhD student at GSK Vaccines and INRA has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° 607611, during the conduct of the study.

Studies described at Chapters III, IV and V are scientific data in publication process. Chapter III was submitted to Journal of Proteomics.