Collagen-like sequences as adhesion factors encoded by the pathogenic bacterium *Burkholderia cenocepacia*: computational analysis and experimental studies

Ricardo Nuno Marques Estevens

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisor(s): Prof. Arsénio do Carmo Sales Mendes Fialho

Dr. Dalila Madeira Nascimento Mil-homens

Examination Committee

Chairperson: Prof. Leonilde de Fátima Morais Moreira

Supervisor: Prof. Arsénio do Carmo Sales Mendes Fialho

Member(s) of the Committee: Dr. Sandra Viegas

November 2022

Preface

The work presented in this thesis was performed at the Institute for Bioengineering and Biosciences of Instituto Superior Técnico (Lisbon, Portugal), during the period September 2021 - July 2022, under the supervision of Prof. Arsénio Fialho and Dr. Dalila Mil-Homens

This work was supported by national funds from FCT - Fundação para Ciência e a Tecnologia, I.P., in the scope of the project UIDB/04565/2020 and UIDP/04565/2020 of the Research Unit Institute for Bioengineering and Biosciences - iBB and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy - i4HB.

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa

Acknowledgements

To begin, I would like to thank Professor Arsénio Fialho for allowing me to do this work with him, and for the corrections, advice, opportunities, and patience. Next, I would like to thank Dalila for her spectacular accompaniment throughout my time in the laboratory and beyond. Thank you very much for your advice and patience, and I hope I have not been too annoying. I would also like to thank all the people in the laboratory who were always very helpful.

Next, I would like to thank my friends, many of them in the same struggle to finish the thesis, for their support. A special thanks to my friend Iris for your computer, I would not finish this without it. To my dear friends at CIRL, thank you for changing the temperature!

Finally, thank you to my parents and my brother, always there. And to my girlfriend, who always supported me. Thank you very much!

Abstract

Burkholderia cenocepacia is a multi-drug resistant pathogen capable of causing chronic infections in cystic fibrosis or immunocompromised patients. This species synthesizes a panoply of adhesins that are critical for adhesion to the host cell. A transcriptomic study revealed that during the first contact of *B. cenocepacia* K56-2 to giant plasma membrane vesicles derived from human bronchial cells, a group of “collagen-like genes” are overrepresented among the induced genes. Since the role in adhesion of these genes in *B. cenocepacia* is unexplored, in this study we provide a bioinformatic analysis. We identified 75 collagen-like proteins (CLPs) containing the Bacterial Collagen Middle Region (Col_mid_reg) in Burkholderia cepacia complex (Bcc) members, analyzed its phylogenetic distribution, identified that CLPs are formed by extensive intrinsically disordered regions, and discussed how these regions may increase the efficiency of CLPs as adhesion factors. Next, we analyzed 5 CLPs paralogs in *B. cenocepacia* J2315. Additionally, the functional analysis of gene *bcal1524* encoding a CLP was carried out. A mutant of *B. cenocepacia* K56-2 was constructed and a phenotypic evaluation was performed regarding biofilm formation, adhesion to host cells (bronchial cell line), and to components of the extracellular matrix (ECM), motility, and virulence using the model *Galleria mellonella*. Our data suggest that the absence of this protein decreases the motility capacity and increases the adherence to host cells. However, no differences were observed in biofilm formation, virulence or adhesion to different components of the ECM. Collectively, our results suggest that the negligible effects observed may represent functional redundancy and compensation effects of the paralogous copies. More studies are necessary to understand the role of CLPs in this species.

Keywords: *B. cenocepacia*; cell adhesion; bacterial collagen-like proteins; low complexity regions; intrinsic disorder proteins

Resumo

Burkholderia cenocepacia é um patógeno multi-resistente capaz de causar infeções crónicas em doentes com fibrose cística ou imunocomprometidos. Esta espécie sintetiza uma panóplia de adesinas que são essenciais para a adesão a células hospedeiras. Uma análise transcritómica revelou que durante o primeiro contacto de *B. cenocepacia* K56-2 com vesículas gigantes de membrana plasmática derivadas de células bronquiais humanas, um grupo de "genes tipo colagénio" estão sobre-representados entre os genes induzidos. Uma vez que o papel destes genes na adesão em *B. cenocepacia* não é explorado, neste estudo fornecemos uma análise bioinformática. Identificámos 75 proteínas semelhantes ao colagénio (CLPs) contendo a *Bacterial Collagen Middle Region* (Col_mid_reg) nos membros do complexo Burkholderia cepacia (Bcc), analisámos a sua distribuição filogenética, identificámos que as CLPs são formadas por extensas regiões intrinsecamente desordenadas, e discutimos como estas regiões podem aumentar a eficiência das CLPs como fatores de adesão. Analisámos também 5 CLPs parálogos em *B. cenocepacia* J2315. Além disso, foi realizada a análise funcional do gene *bcal1524* que codifica uma CLP. Foi construído um mutante de *B. cenocepacia* K56-2 e foi realizada uma avaliação fenotípica relativamente à formação de biofilme, adesão às células hospedeiras (linha de células bronquiais), e aos componentes da matriz extracelular (ECM), motilidade, e virulência utilizando o modelo *Galleria mellonella*. No total, os nossos dados sugerem que a ausência desta proteína diminui a capacidade de motilidade e aumenta a aderência às células hospedeiras. Contudo, não foram observadas diferenças na formação do biofilme, virulência ou adesão a diferentes componentes do ECM. Coletivamente, os nossos resultados sugerem que os efeitos negligenciáveis observados podem representar redundância funcional e efeitos de compensação dos genes parálogos. São necessários mais estudos para compreender o papel das CLPs nesta espécie.

Palavras-chave: *B. cenocepacia*; adesão celular; proteínas tipo colagénio bacterianas; regiões de baixa complexidade; proteínas intrinsecamente desordenadas

Table of contents

List of Figures

List of Tables

1. Introduction

1.1. Bacterial adhesion: the ability to stick

Bacteria are one of the most diverse organisms known, able to colonize almost all habitats. Capable of establishing populations both in liquid and solid environments, most bacteria live in a sessile form, attached to a surface, from natural of artificial origin[1–6]. These different substrates possess diverse physical and chemical properties, indicating that bacteria have great adaptation capabilities. Sessile microbial populations are more common than planktonic ones because the adhesion to a surface comes with a series of advantages to the bacterial cell itself or the bacterial population: from a higher concentration of nutrients, passing by protection against predators or the environment[7], and, in the case of pathogens, against immune responses of the host organism[2].

The event of bacterial adhesion is a very complex mechanism not fully understood, dependent on several factors, such as the nature of the surface, the surrounding environment, and the cell itself[8]. Adherence to an abiotic surface or a living surface (for example, other prokaryotic or eukaryotic cells/tissues) occurs in different manners (Figure 1). Non-specific interactions are usually associated with adhesion to inert surfaces, while the interaction with biological surfaces is performed, at least in part, through the establishment of specific ligand-receptor interactions[9]. However, before the contact between bacteria and the substrate occurs, it is important to focus on the conditioning layer phenomenon[10]. The organic and inorganic molecules present in the bulk flow will accumulate on the surface, increasing their concentration near the solid-liquid interface[11]. This deposit of molecules such as nutrients and proteins seems to have an important role in the first stages of bacterial adhesion since it can alter the physical and chemical properties of the substratum[12]. The conditioning film created can have a positive and a negative impact on the adhesion efficiency, since the composition of the film will depend on the substances diluted in the bulk[10]. Thus, some substances may reduce bacterial ability to adhere, while others may facilitate the interactions with the bacteria, increasing their adherence capacity[13] (Figure 1. Bacteria can adhere to different surfaces, like ship hulls, medical devices, or the human body. Adhesion is influenced by different factors, from the nature of the surface to the bacterial surface composition, passing by the formation of a conditioning layer. Figure created with biorender.com.).

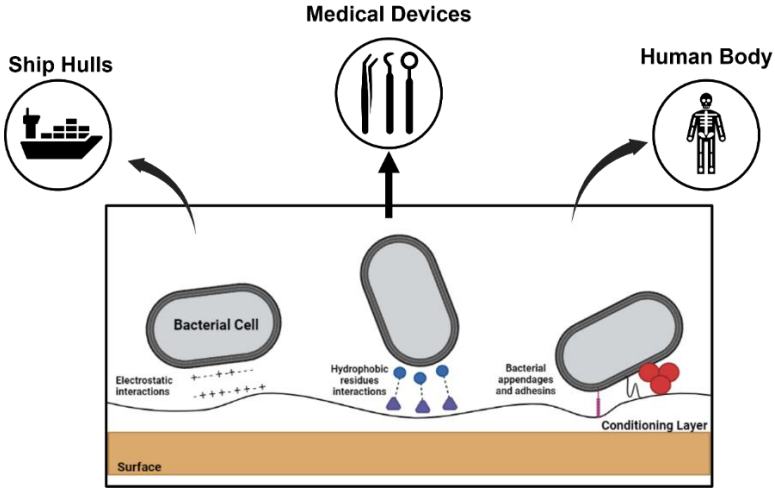


Figure 1. Bacteria can adhere to different surfaces, like ship hulls, medical devices, or the human body. Adhesion is influenced by different factors, from the nature of the surface to the bacterial surface composition, passing by the formation of a conditioning layer. Figure created with biorender.com.

Focusing on the adhesion process of a bacterium to a surface, and describing it in a simplified way, the process can be divided into two phases[11,14]. The first step requires the bacterium to get close to the surface, either by random movement (bacteria can be dragged by the fluid flow) or by bacteria-directed movements using appendages such as pili and flagella[15]. When the bacterium is close enough to the substratum, the attachment is dependent on the balance between the attractive and repulsive forces involved[16]. If the attractive forces outweigh the repulsive ones, the cell contacts the surface. The most common forces involved in these interactions are van der Waals forces, long-ranged, usually attractive, and the strongest when the cell is near to the solid surface, leading to the adherence of the bacterium[8]; electrostatic interactions, that are dependent not only on the pH and ionic strength of the liquid environment but also on Brownian motion forces; and hydrophobic interactions, which can be both attractive and repulsive, depending on the surface characteristics, the bacterium, and the surrounding environment[16]. Additionally, steric forces are in certain cases key to successful adhesion. Different species of bacteria present complex and diversified outer membranes, with varied compositions and charges. This layer can generate steric forces also involved in the adhesion process[17]. Moreover, the characteristics of the surface itself are determining factors for weak or strong adhesion. The simple topography of the substratum on a micrometre scale level can influence its hydrophobic properties, an important parameter for bacterial adhesion, as mentioned before. Irregular surfaces facilitate adherence due to an increased surface area that will maximize the contact between the cell and the surface and reduce the shear stress to which attached cells are subjected[13,18]. Besides the physical characteristics, the chemical properties of the solid surfaces are also an important factor that influences bacterial adhesion, being a parameter usually addressed when trying to impede the establishment of growing populations of bacteria. Since bacteria usually possess a negatively charged cell wall[8], even if their composition and chemical properties are extremely variable, surfaces composed of neutral or positively charged materials are more easily colonized by the cells[19]. The hydrophobicity of the materials is a relevant factor, as well. More hydrophobic surfaces will preferentially interact with bacteria whose cell envelope is more hydrophobic, and the opposite will also occur[19]. During this first contact, bacteria can still be easily removed from the surface[14]. The next step, which will lead to an irreversible attachment

from the cells, is mediated by specific ligands from the bacterial cell wall that will be produced and complex with the surface. These moieties that will interact with the surface can be of different chemical identities: lipids, proteins, or sugars. When this step is completed, the removal of bacteria requires stronger physical or chemical agents. The bacterial cell envelope, composed of various proteins, lipids, exopolysaccharides, and fimbrial and non-fimbrial structures, presents such complexity and heterogeneity that makes it difficult to mention and describe all the different molecular tools used by these organisms to adhere to and establish a population[20].

1.2. Methods for measuring bacterial adhesion

For a better understanding of this phenomenon, researchers developed through the years a multitude of different techniques and assays to measure the adhesion force to surfaces[21]. These techniques may be qualitative (percentage of cells that remain attached) or quantitative (magnitude of adhesion forces). For a more general review, addressing several assays, consult Alam et al[21]. As an example of a qualitative technique, in the Centrifugation Assay, the centrifugal shear force is applied to the cells, measuring the adherent forces[22]. The substrate containing adherent cells is centrifuged at a certain speed, applying detachment force. The plate with the adherent cells can also be inverted, combining the centrifugal with the gravitational force. In this case, non-adherent cells will be lost (Figure 2). The results obtained are expressed as the shear force necessary to detach 50% of the cells[21]. Since this technique does not allow the distinction between living cells from debris or dead cells, it only provides a qualitative metric[21].

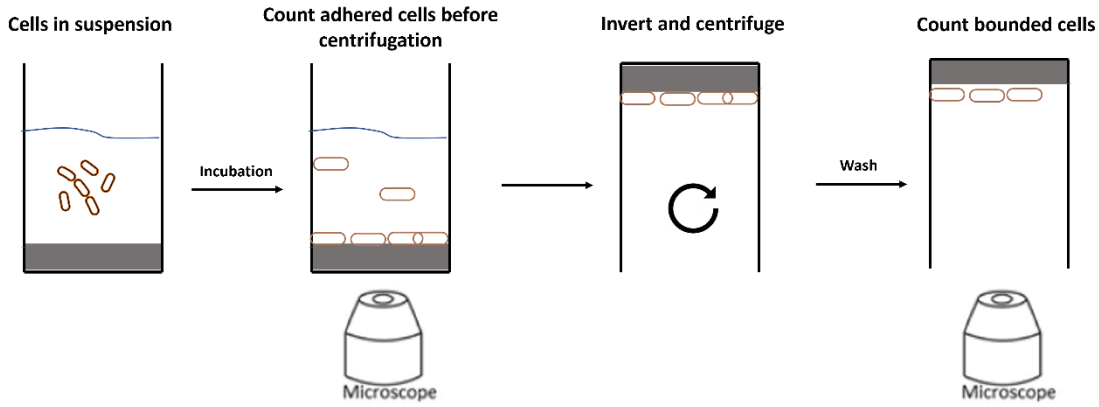


Figure 2. Schematic of the Centrifugation Assay. Brief description of the steps needed for the measurement of the adhesion forces involved. Figure created with biorender.com.

Atomic force microscopy (AFM), on the other hand, is a quantitative method. With this technique, researchers can measure single-cell interactions between the bacterium and the substrate or other cells[23]. By using the AFM-cantilevers functionalized with a single bacterium, it is possible to study specific adhesion, as Beaussart and colleagues demonstrated by measuring glycopolymer-lectin interactions in *Lactobacillus plantarum*[24]. Additionally, this technique allows the measurement of nonspecific adhesion forces. Thewes and colleagues studied the unspecific forces involved in the adherence mechanism of *Staphylococcus carnosus*[23]. In this study, the authors were able to demonstrate that the nonspecific adhesion of *S. carnosus* is influenced by van der Waals, electrostatic

and hydrophobic forces. Single-molecule atomic force microscopy was used, as an example, to study the binding mechanism of a specific adhesin of *Burkholderia cenocepacia*[25]. AFM allows several approaches to the adhesion problem, being capable of measuring a single molecule as well as complete cell influence in the physiological environment[21].

1.3. Methods for identifying bacterial adhesion molecules

Identifying the surface components of a bacteria responsible for adhesion, like proteins, is of extreme importance since many of these structures are involved in pathogenicity and their understanding may lead to the creation of novel therapies[26]. For that reason, the development of techniques appropriate for the study of the "adhesiome" of bacteria is needed. In the initial phase, in silico analysis are very important and allows the identification and screening of multiple genes or proteins of interest, through the prediction of structure and function[27,28]. Additionally, transcriptomic analysis is also an important method to identify genes that may be related to the adhesion process. Either by analysing alteration in gene expression during adhesion assays or by analysing the behaviour of identified adhesin encoding genes to eventual therapeutic agents[29,30]. Methods using gel-free enzymatic digestions of the total cellular adhesiome in parallel with Mass Spectroscopy (MS), for example, are also extremely important[26]. Another useful technique is the blot rolling assay, a method which allows the identification of adhesins under shear stress[31]. Regarding the aim of studying proteins associated with immunoreactivity, the usage of antibody affinity magnetic beads allows the separation of specific proteins, following enzymatic digestion and MS for antigen identification. This method also allows the examination of direct or indirect antigen-associated complexes. Associating these separation and identification techniques with MS and bioinformatic tools is important for the investigation of surface proteins and understanding their structures and functions, namely for adhesion and pathogenicity-related processes.

1.4. The major classes of bacterial adhesins

1.4.1. Flagella and fimbrial adhesins

Flagella, a bacterial filamentous appendage commonly associated with movement, can also play an important role in adhesion[32] (Figure 3). Besides the significance of driving the cell to the proximity of the substratum, the flagella are responsible for the initial surface sensing during the adherence process[33] and for adhering to determined targets[32]. Girón et. al[34], for example, showed flagella importance in the adherence of pathogenic *Escherichia coli* to epithelial cells. Flagella are also described as being independent of the substrate, unravelling their role in the initial unspecific adhesion[35]. Fimbrial adhesins and pili are described as essential in the infection process in bacteria[36], with its impact studied in several bacterial species, like *Klebsiella penumoniae*[37], *E. coli*[38] or *Salmonella*[39].

Apart from flagella, other filamentous structures present on the bacteria surface possess a determinant role in the adhesion mechanism. Pili, or fimbrial adhesins, are complex polymeric fibres, composed of repetitions of protein subunits arranged into a helix, creating a long hair-like nanofiber[40] (Figure 3). Each bacterial cell usually possesses various types of pili that are present on the cell surface and, among

other functions, are involved in cell adherence to abiotic and biotic surfaces[20]. Although most of the fimbria known can establish nonspecific bindings to the substrates[8], some molecules comprise an adhesion factor in the extremity, conferring specificity to the adhesin[5]. For example, the study performed by Pratt and colleagues[41] showed the role of type I Pili in the initial steps of adherence of Gram-negative *E. coli* and the existence of a mannose-specific adhesin, FimH that facilitates the interaction between the bacterial cell and abiotic surfaces. In Gram-positive bacteria, pili have gone unnoticed until recently, mainly due to reduced dimensions[42]. However, pili have been observed, among other groups, in *Corynebacteria* and *Streptococci*[40]. In fact, in *Streptococcus agalactiae*, it was already proven the role of PI-2a pilus in the adherence to human lung epithelial cells[43].

1.4.2. Non-fimbrial Adhesins

Besides the elongated forms like fimbriae, bacterial cells also present a multitude of nonpolymeric structures, non-fimbrial adhesins. This type of adhesins is described as involved in the attachment process to both abiotic surfaces and host cells/tissues. Due to their usually reduced size, these proteins are related to close contact with the substrate[20]. Non fimbrial adhesins are widely distributed in Gram-negative bacteria, and besides their importance for biofilm formation[44], they are capable of recognizing different elements of the surface for binding, including elements of the extracellular matrix of eukaryotic cells, like collagen and fibronectin[5]. Fibronectin binding proteins are identified in Gram-positive bacteria, for example, in *Staphylococcus aureus*, with an array of functions related to binding to fibrinogen, to plasminogen and promoting biofilm formation[45]. Glycoproteins are also important for the adhesion process, due to the facilitation role of glycans in the proteins[46] (Figure 3). In bacteria, the glycosylated proteins are biosynthesized by two different mechanisms and their function is being studied in more and more organisms, like *B. cenocepacia, Francisella tularensis* or *E. coli*, with functions as diversified as evasion of the immune system of the host or involvement in diffuse adherence[47–50].

1.4.3. Autotransporter adhesins

Autotransporter adhesins (ATADs), monomeric or oligomeric nanofibers, are nonfimbrial adhesins present in Gram-negative bacteria (Figure 3). This type of adhesin is denominated “autotransporter” due to the mechanism by which the protein is translocated, which is independent of any known form of cytosolic energy source[51]. This mechanism known as the Type V secretion system is very common in Gram-negative bacteria and is used by the cells to export structures with diversified functions, from adhesins to toxic proteins[52]. The autotransporter adhesins are responsible for interaction with both biotic and abiotic surfaces. One example of these ATADs is AIDA-I, an adhesin from *E. coli*[53] that confer diffuse adhesion to different cell types. Diffuse adhesion is the term used when the bacteria adhere uniformly to a surface. This ability occurs due to the glycosylation of AIDA-I and its receptors, scattered across the cell surface[54]. Other adhesins are not so general as AIDA-I, targeting more specific receptors in the host cells or surfaces. FHA, from *Bordetella pertussis*, favours specific carbohydrates, binding specifically to surface glycans in human lung epithelial cells, as shown by Tuomanen and colleagues[55]. Another subfamily of autotransporters was also discovered: the trimeric autotransporter adhesin (TAAs)[56]. The difference between this family of ATADs is the existence of a trimeric translocator domain,

in which three subunits associate to form the pore for translocation of the also trimeric protein[56]. Trimeric autotransporters are usually constituted by several repetitive sequences and structures, which seems to lead to the existence of more than one binding site, increasing the propensity to surfaces presenting the respective receptor in high quantities[57]. The most studied example of this structure is the YadA adhesin from *Yersinia enterocolitica* that confers to the cell the ability to bind to collagen, fibronectin, and laminin[58], among other characteristics important for the pathogenicity of *Y. enterocolitica*. So far, the TAAs reported present an adhesion function, even though their tertiary and quaternary structure can be very diversified[40]. This type of adhesins is also well represented in *Burkholderia cenocepacia*, possessing a crucial role in several stages of infection, being important for biofilm formation, motility, immune system evasion, adhesion and invasion to host cells[59–62]. These results highlight the versatility and importance of TAAs for the adhesion process.

1.5. Bacterial collagen-like proteins (CLPs): expanding the adhesion portfolio

Contrarily to eukaryotic collagen, with its physical and chemical properties explored and researched immensely, bacterial collagen caught the interest of the scientific community more recently. Metazoan collagen plays important roles in eukaryotic organisms, being associated with tissue structure, extracellular matrix composition and function, and signal transduction processes[63]. Existing various types of metazoan collagen, all share one structural feature: a triple helical structure, composed of three chains with a characteristic pattern of Gly-Xx-Yy, with Xx and Yy being proline and hydroxyproline, respectively[64,65]. In bacteria, the Gly-Xx-Yy pattern has been identified in various protein databases[66,67]. From the ones that were possible to express *in vitro* and confirm the triple helix structure, it was possible to observe the lack of hydroxyproline residues in the chains[68,69], contrarily to its frequency in eukaryotic collagen. In bacteria, besides the existence of these GXY domains, a specific domain designated Bacterial Collagen Middle Region (PF15984 in the Interpro tool) has been identified. This domain is not present in all the proteins possessing GXY repeats, and not all proteins possessing this domain present GXY repeats. This indicates the possible existence of two different types of CLPs in bacteria: a type I, containing GXY repeats capable of originating triple helixes; and a type II, containing the Bacterial Collagen Middle Region (Figure 3).

Regarding the origin of the collagen-like sequences in bacteria, two options were hypothesized: horizontal transfer of collagenous sequences between eukaryotes and bacteria[66,70], or the emergence of collagen repeats in bacteria *de novo*, resulting from spontaneous mutations, duplication of repetitive sequences, or domain organization[71,72]. The *de novo* appearance of these sequences in prokaryotes may suggest convergent evolution between prokaryotes and eukaryotes[72,73]. Considering the molecular architecture of this type of protein, and after analyzing and comparing it with the eukaryotic collagen, it is considered that the prokaryotic collagenome is more heterogeneous in terms of length and sequence composition than the eukaryotic collagenome[74].

CLPs possessing the GXY triplet repeats were studied in pathogenic bacteria regarding their structure and function. Studies with different bacterial species found that these proteins are seemingly associated with the pathogenic process, playing an important role in cell adhesion, invasion, or immune evasion[75,76].

Focusing on the adhesion role, the most well-studied CLPs so far are present in *Streptococci* strains, being present in several pathogenic ones, like *S. pyogenes,* or *S. pneumoniae*[77]. One of these proteins is Scl1, the first CLP mentioned in *S. pyogenes*[78,79]. Using bioinformatics tools, it was possible to observe that Scl1 was cell wall anchored, homotrimeric, and lollipop-like structured. This protein possesses a collagen-like rod-shaped domain (CLD) that projects the noncollagenous V domain away from the cell surface[71]. This V domain was proven to have binding specificities to plasma lipoproteins[68]. Additionally, Chen and colleagues[76] proved the role of Scl1 in the adhesion of *S. pyogenes* to epithelial cells, showing also that this adhesion process is mediated by interactions with protein receptors α2 and β1 integrins on the surface of the epithelial cells.

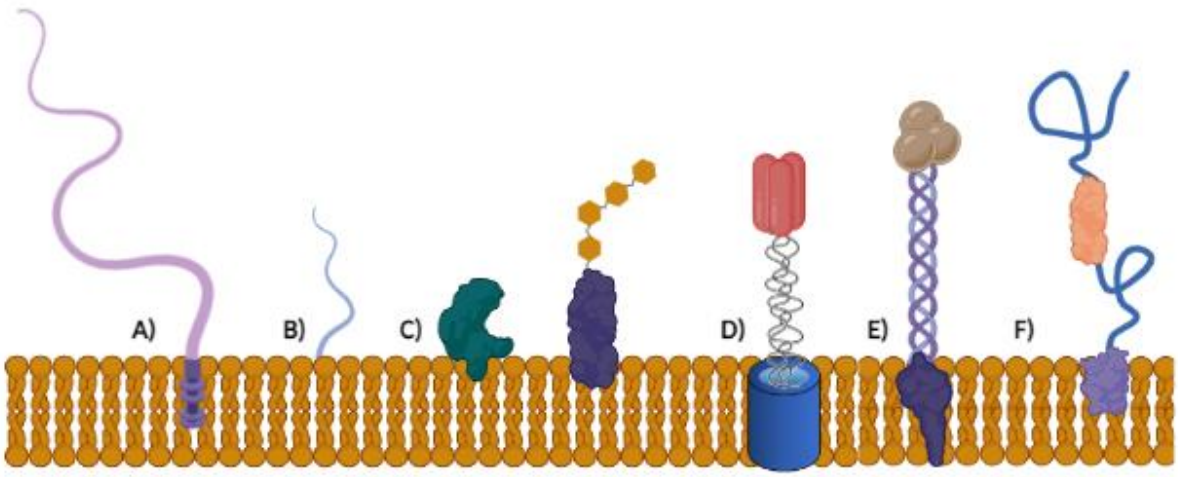


Figure 3. Representation of different classes of adhesins. **A)** Flagellum; **B)** Pili; **C)** Non-fimbrial adhesins, globular and glycoproteins; **D)** Trimeric Autotransporter Adhesin; **E)** Collagen-like protein type I; **F)** Collagen-like protein type II. Figure created with biorender.com

In *Burkholderia pseudomallei*, a human pathogen, several putative CLPs (Bucl proteins containing the GXY repeats, IPR008160 in the InterPro tool) were identified[80]. The authors registered the presence of 13 genes in *B. pseudomallei*, observing genomic rearrangements in the *bucl loci* between strains[80]. The 13 proteins possessed a collagen-like region flanked by N- and C- noncollagenous terminal regions. The N- and C- terminals of the proteins were conserved among three different *Burkholderia* species, while the collagen-like region varied significantly between species. Bachert and colleagues predicted the structure of those proteins and identified the presence of putative domains known for their role in pathogenicity. Bucl8, for example, was predicted to possess an outer membrane flux (OEP) domain[80], known for allowing bacteria to export compounds to the exterior, including anti-microbial agents. Bucl8 structure was assessed by Grund and colleagues, who observed that this protein is a homotrimeric molecule, possessing the CL region. This region is characteristic of its (GAS) pattern, although the number of repeats is variable among strains[81]. The authors also observed the importance of the CL domain in the ability of the protein to bind to fibronectin. Moreover, the authors concluded that Bucl8 is part of an efflux system involved in antibiotic resistance[81], demonstrating the importance of this collagen-like protein in the pathogenicity of *B. pseudomallei*.

As described above, CLPs well-characterized are associated with the cell surface[82]. Proteins identified in various members of *Bacillus* and *Clostridium* contained phage-related and LPTX anchor motifs directing proteins to the cell. Moreover, conservative domains with ligand-binding functions were observed in CLPs found in extremophile organisms, with even more proteins registered with unknown functions[82].

Bacterial collagen was already identified as having high biotechnological potential, with applications in the cosmetics and biomedicine fields[82]. Being non-toxic, nonimmunogenic, and thermally stable at body temperature without the need for secondary modifications[69], recombinant proteins possess high potential. Additionally, the bacterial original presents an advantage consistent with the animal-free trend in consumables worldwide[69]. Besides, recombinant CLPs can be produced in *E. coli* by fermentation in bioreactors with a higher yield than the production of animal collagen[83], although the risk of endotoxin contamination exists[69]. The ease by which biologically active sequences and domains can be added to bacterial collagen is also seen as an attractive characteristic of this system. Bacterial CLPs may be used to produce scaffolds, gels, or filaments that can be used for drug delivery and tissue regeneration, for example[69]. To this date, protein Scl2 of *S. pyogenes* is one of the best-studied collagenous proteins regarding biotechnological applications. Containing no biological activity by itself, recombination with active domains leads to the production of hydrogels used to enhance osteoblast adhesion and spreading on titanium materials[84] or used in microspheres for wound healing[85]. However, more studies are still needed to fully understand the potential of CLPs, making this an emerging area of research.

1.6. Type II CLPs: bringing disorder to facilitate adhesion?

Type II CLPs are predicted to contain great extensions of intrinsically disordered segments. Intrinsically Disrodered Proteins (IDPs) are defined as lacking stable secondary and/or tertiary structures. Disordered proteins possess a different composition in terms of amino acids when compared to structured proteins[86]. This type of protein is enriched in amino acids like proline and glycine, for example, amino acids characterized as being charged and structure destabilizing[86]. Moreover, the frequency of hydrophobic and aromatic residues is diminished in these proteins, diminishing the ordered regions[87]. It is to be noted, also, that IDPs usually possess low complexity regions, characterized as regions with little diversity in amino acid composition and repetitiveness[88]. These LCRs are identified as being possibly located in the extremities of the protein, with these domains possessing biological roles[88]. Initially, these proteins were thought of as being uncommon and unspecific, but they are now identified in all kingdoms, being involved in several biological functions[89]. In the case of bacteria, it is estimated that the percentage of IDPs or proteins with disordered regions varies from 18 to 28%, a relatively small amount compared with eukaryotic organisms[90]. In terms of function, for bacteria, Peng and colleagues identified a correlation between intrinsically disordered regions and DNA and RNA binding, as well as sporulation, catabolic and metabolic processes, and pathogenesis[91]. Mier and colleagues suggest, also, that low complexity regions are more represented in the outer membrane and extracellular proteins in bacteria, being involved in host-pathogen interactions[92]. These discoveries unravel the importance of disorder and low complexity regions in proteins associated with pathogenicity and adhesion processes, encouraging further study of the importance of these domains in bacterial proteins.

1.7. Adhesion as a target for anti-microbial therapies

Adhesion molecules, generally called “Adhesins”, are present in many pathogens, being of extreme importance for their capacity for infection and virulence. Considering that bacterial infections are a major cause of morbidity and mortality worldwide and the rising problem of antibiotic-resistant bacteria, new mechanisms should be addressed to prevent and treat bacterial infections. Since the adhesion of bacteria to surfaces, like human tissues or cells, is vital for their ability to colonize and infect the host, interfering with this process is a potential target for treatment and prevention of infections[93].

The adhesion mechanism is a multivalent process. Different steps require different agents, from the initial adsorption, mediated by non-specific interactions, to the high-affinity binding between bacterial adhesins and surface receptors. Different steps can be potentially targeted for anti-adhesion therapies during this multi-phased process. From impairing the initial non-specific interactions by altering surfaces’ chemical or physical, to inhibiting the biogenesis of adhesins or host receptors[94], passing by developing anti-adhesion compounds that can mimic the bacterial components, impeding their interaction with the host cell

Some antibiotics, namely fluoroquinolone, ciprofloxacin, and amikacin were described as capable of altering the physicochemical properties of the bacterial cell surface. By interfering with protein synthesis, leading to aberrant protein display in the outer membrane, these compounds result in changes in the surface charge of the bacteria, as well as impairing specific interaction with host receptors promoted by defective proteins[94].

Besides influencing the pathogen’s membranes and its adhesins, the depletion of host receptors was also proposed as a strategy to deal with bacterial infections[95]. Svensson and colleagues[96] observed that the inhibition of the enzyme ceramide-specific glycosyltransferase led to a decrease in bacterial colonization in human cells. This happens because that enzyme belongs to the glycosphingolipids (GSLs) biosynthetic pathway, a common host receptor for bacterial adhesins.

Carbohydrates are also commonly used by bacteria to establish interaction with host cells, being present in both the bacteria and the host surface. For that reason, researchers focused on the development and usage of synthetic glycosides that would compete with the bacterial adhesins, limiting their ability to interact with the host carbohydrates. Several compounds were developed through the years (see Sharon[97] for a review), with different targets, but the uropathogenic *E. coli* adhesin FimH is one of the most studied. For this adhesin, researchers developed multivalent inhibitors, taking into account the affinity of FimH to mannosylated structures on the host cell surface. Coupling FimH antagonists with multivalent scaffolds, potent anti-adhesives were developed[98]. Moreover, the same reasoning of competitive inhibition is applied to the development of peptide-based compounds. However, the *in vitro* potential[99] was not fulfilled yet.

Finally, the development of anti-adhesive agents also includes the development of anti-adhesion antibodies and vaccines. Different approaches are possible, from immunizing the host using a bacterial adhesin, a subunit, or an immunogenic fragment. Additionally, a DNA vaccine that encodes the adhesin or an immunogenic part can be administered to the host[93]. *Salmonella enterica* serovar Typhi, the cause

of typhoid fever, contains in its membrane the adhesin T2544, a potential target for vaccine development[100]. Ghosh et al showed that infection of mice with T2544 triggered a strong immunogenic response, with T2544-specific antibodies detectable 4 months after immunization. The authors also showed that mice immunized were resistant to further bacterial infection[100], proving that this adhesin might be a good candidate for future vaccine development. However, the typical multifactorial adhesion of bacterial pathogens complicates the process. T2455-immunized subjects were not completely protected against the disease since other structures from the bacteria promote cell adhesion and infection[100]. This is, in fact, one of the weak points of anti-adhesion therapies. Adhesion of bacteria is ensured by several adhesins, which may weaken the activity of a vaccine or antibody for a specific adhesin[93]. On the other hand, many adhesins possess a high degree of conservation, turning them into potential vaccine targets[101].

As presented briefly in this work, anti-adhesion strategies are a major area of research as an alternative to other antimicrobial approaches for fighting microbial resistance. This is because anti-adhesive strategies do not interfere with bacterial viability, exerting any selective pressure on the pathogen[93]. Furthermore, anti-adhesive compounds are stable under physiological conditions. Since anti-adhesion compounds are designed to mimic and compete with bacterial or host structures, they will be less susceptible to degradation than other antimicrobial agents with other modes of action[93].

Targeting adhesion for antimicrobial agents presents itself as a promising strategy to relieve pressure on antibiotic resistance[102]. Impairing bacterial ability to adhere will leave bacteria in a planktonic state, susceptible to removal by physical or immune action unless the proliferation rate is superior to the removal rate[103]. It is important to acknowledge that anti-adhesion therapies will impede bacterial adhesion but will not impede their proliferation or clearance[104].

1.8. The human respiratory pathogen *Burkholderia cenocepacia*: a model organism to study cell adhesion

Burkholderia cenocepacia (*B. cenocepacia*) is a member species of a defined group of gram-negative bacteria named The *Burkholderia cepacia* complex (Bcc)[105]. Several members of this complex are found in the natural environments[106], such as soil or aquatic habitats, participating in commensal interactions with plants[105] and even degrading pollutants like trichloroethylene[107]. Besides that, members of the Bcc complex have emerged as problematic opportunistic pathogens in immunocompromised and cystic fibrosis (CF) patients[108,109]. Through genomic plasticity, they appear to be highly adapted to the respiratory tract and very prone to develop antibiotic resistance. CF patients are especially vulnerable to lung infections caused by these pathogens. Isles et al[109] observed that CF patients suffer a rapid clinical deterioration when infected with a Bcc strain, developing, in some cases, necrotizing pneumonia and septicaemia, resulting in death. Contacts between individuals or contaminated surfaces are the main transmission ways of Bcc infection since it is thought to occur via aerosol droplets or physical contact[106]. Due to the ease of transmission, medical organizations worldwide write recommendations to avoid CF outbreaks. Bcc infected patients are recommended to stay separated from each other when

in clinical facilities. Contact with CF patients outside the hospital environment is also discouraged for infected patients, affecting the individual's well-being[106].

Burkholderia cenocepacia and *Burkholderia multivorans,* have been identified as the Bcc members more prevalent in CF infections[110]. However, most of the cases of transmission are caused by *B. cenocepacia*[111]. From this species, three specific strains led to three epidemic events that generated significant mortality and morbidity in CF patients over the world: ET-12, Midwest, and PHDC lineages[110]. *B. cenocepacia* ET-12 strain was responsible for a large epidemic in Canada and United Kingdom, making it the most studied strain of this species[112].

B. cenocepacia possesses a large repertoire of virulence factors, including adhesins, invasins, enzymes, toxins, secretion, and quorum sensing systems[113,114]. Focusing on the adhesion process, several studies demonstrated the variety and complexity of the components. From flagella[115] to outer membrane proteins[116], passing by lipoproteins[117], several molecules have been documented as participating in the adhesion mechanism in *B. cenocepacia*. Trimeric autotransporter adhesins are one of the most well-studied adhesins in Bcc species, to date. Mil-Homens and colleagues[59] identified an adhesion cluster in the genome of *B. cenocepacia* J2315, studying the BCAM0224 protein and observing its importance in cellular adhesion and virulence since this protein represents a new collagen-binding trimeric autotransporter adhesin. Other TAAs were studied in detail, proving their importance for *B. cenocepacia* infection ability, being important not only for adhesion but also for biofilm formation, motility, and invasion of host cells[61,62,118,119]. Besides TAAs, cable pili and their associated adhesins are also well documented for their role in adherence and virulence[120].

Bearing that in mind, and as the initial interaction between a bacterium and the host cell surface is of crucial importance[5], Pimenta et al performed an RNA-sequencing and the transcriptome of *B. cenocepacia* K56-2 was analyzed after the first contact with host cell-derived vesicles[29]. This work allowed the identification of various genes involved in the bacteria's adaptation to an adherence situation. The authors observed that the interaction of the bacteria with the bronchial cell membrane altered the metabolism of the bacteria. Genes involved in pathways like the TCA cycle and glycolysis were downregulated, while genes involved in sulphur and nitrogen metabolism, for example, were upregulated. These observations indicate the capacity of bacteria to adapt to different environments. Regarding adherence, in this study, the authors observed that several membrane structures important in bacterial adherence and virulence had their genes overexpressed during the initial phase of adhesion. Genes encoding components of Flp type pilus, bapA cable pilus associated adhesins, and the *BCAM2418* and *BCAS0236* trimeric autotransporter adhesins were identified, revealing their role in the initial stage of adhesion of *B. cenocepacia* in the tested conditions[29]. However, the authors identified many down- or upregulated genes that were not been studied yet. Some of those genes are thought to be lipoproteins, probably associated with the virulence of this species. Three of those genes are the *bcal1523, bcal1524* and the *bcam0695*, which were registered as being overexpressed 98, 138 and 131-fold, respectively[29]. Further investigation revealed that these genes are identified as putative collagen-like genes and since they were detected with enhanced expression during the initial stages of adhesion of *B. cenocepacia* to cell-derived vesicles, it is hypothesized that their role is related to the

adhesion process. Knowing that, this work aims to study this class of proteins, performing computational analysis of the collagen-like proteins present in the Bcc and in *B. cenocepacia*, specifically. Besides, we aim to understand the importance of the BCAM1524 protein in the adhesion process and virulence. For that, a deletion mutant of *B. cenocepacia* was constructed and a phenotypic analysis versus the wild-type strain was performed. Several tests were conducted to evaluate differences in growth, motility, biofilm formation, adhesion to bronchial pulmonary host cells (16HBE14o- cell line), and components of the extracellular matrix (collagen type I and IV and fibronectin). Additionally, two truncates were constructed for the BCAL1524 protein to continue the evaluation of the protein's domains and their function. We intend to understand the importance of the collagen-like proteins in the adhesion process of the pathogenic organism *B. cenocepacia* as well as increase the information and knowledge regarding bacterial collagen and its function.

2. Materials and Methods

2.1. Computational Analysis

The collagen-like proteins were identified by searching for sequences possessing the Bacterial Collagen Middle Region (Collagen_mid) family (PF15984) in one strain representative of each species of the Bcc. Protein sequences were obtained using the Burkholderia Genome Online Database (https://www.burkholderia.com/)[121], as well as the neighbourhood of the collagen-like encoding genes from *B. cenocepacia* J2315. The subcellular localization of each protein was predicted using PSORTb program (https://www.psort.org/psortb/)[122]. The sequences (from collagen-like proteins and 16SRNA) were used to obtain phylogenetic trees with phylogeny.fr using the "one-click workflow" (MAFFT for multiple alignments, BMGE for alignment curation, PhyML for tree inference and Newick Display for tree rendering)[123]. Specific domains and motifs were searched using the Pfam database (pfam.xfam.org) and InterPro online tool (https://www.ebi.ac.uk/interpro/). To identify the tandem repeats the XSTREAM Web Interface tool was used with default settings[124]. For disordered region prediction, the flDPnn webserver (http://biomine.cs.vcu.edu/servers/flDPnn/)[125] and the SMART webserver (http://smart.embl-heidelberg.de/)[126] was used. The three-dimension (3D) structural prediction of each domain (N-terminal, Collagen-mid and C-terminal) of the *B. cenocepacia* J2315 proteins were performed with the I-TASSER server (https://zhanggroup.org/I-TASSER/)[127].

2.2. Bacterial strains and growth conditions

B. cenocepacia clinical isolate K56-2 (clonally related to *B. cenocepacia* J2315), *E. coli* αDH5, and *E. coli* One Shot® Mach1™-T1^R (Invitrogen, Thermo Fischer Scientific, USA) were used. Bacteria were cultured in Luria-Broth (LB) medium (NZYtech, Portugal) at 37ºC with orbital agitation at 250 rpm. When appropriate, the medium was supplemented with 150 mg/L of ampicillin, 50 mg/L or 100 mg/L of kanamycin (Sigma Aldrich, USA) for *E. coli*, and 100 mg/L, or 150 mg/L of trimethoprim (Sigma Aldrich, USA) for *B. cenocepacia* K56-2 mutants.

2.3. Cell lines and cell culture

An immortalized human bronchial epithelial cell line 16HBE14o-[128] provided by Dr. Grunert (California Pacific Medical Center Research Institute, University of California at San Francisco, USA), was used. Cells were maintained in fibronectin-collagen-coated flasks in a minimum essential medium with Earle's salt (MEM) (Gibco, USA) supplemented with 10% fetal bovine serum (Gibco, USA), 0.292 g/L L-glutamine (Sigma-Aldrich, USA), penicillin (100 U/ml) (Gibco, USA), and streptomycin (100 µg/mL) (Gibco, USA) in a humidified atmosphere at 37 °C with 5% CO_2.

2.4. Attempt to construct *bcam0695* and *bcal1524 B. cenocepacia* K56-2 mutants

The construction of the mutants was performed as described before[119] with some modifications. For the *bcam0695 B. cenocepacia* mutant, a 1086-bp fragment was amplified by PCR with a reaction volume of 25 µL using *Platinum Taq* DNA polymerase (Invitrogen, Thermo Fisher Scientific, USA), DMSO (8%

v/v), $MgCl_2$ (1.5 mM), Buffer (1 x), dNTP (0.02 mM), DNA from *B. cenocepacia* K56-2 (46 ng) and primers at a concentration of 0.2 µM, designed using AmplifX 1.5.4 software: frBCAM0695_fwd' and frBCAM0695_rev (Stabvida, Portugal) (Table 1), containing *BamHI* and *XbaI* restriction sites (underlined), respectively. The PCR program was designed considering the enzyme manufacturer's instructions, with an annealing step at 57ºC and 35 cycles. The amplification was confirmed with a 0.8% agarose gel electrophoresis and the obtained fragment was directly purified from the PCR reaction mixture using the Zymo Research DNA Clean and Concentrator kit (Zymo Research, USA), following the manufacturer's instructions. The plasmid pDrive (Qiagen, Germany) and the purified *bcam0695* fragment were then digested (500 ng of DNA in each digestion) with the restriction enzymes *BamHI* and *XbaI* (Takara, Japan), for 5 hours at 37ºC in a total reaction volume of 20 µL. The pDrive digestion was treated with 1 unit of calf intestinal alkaline phosphatase (CIAP) (Invitrogen, Thermo Fisher Scientific, USA), 15 minutes at 50ºC and 15 minutes at 65ºC, to avoid recircularization. Afterwards, both digestions were purified using the Zymo Research DNA Clean and Concentrator kit (Zymo Research, USA), following the manufacturer's instructions. Finally, the vector and the gene fragment were ligated, with a ratio of 1:3 in a total of 100 ng of DNA. For that, 1 unit of T4 DNA Ligase (Invitrogen, Thermo Fisher Scientific, USA) was used in a total reaction volume of 20 µL. The reaction mixture was kept at room temperature for 30 minutes and then transferred to a thermos for a progressive decrease in temperature and placed at 4ºC overnight. After the ligation, 10 µL of the reaction mixture was used to transform 150 µL of chemically competent *E. coli* αDH5 cells by classical transformation. For this, pre-prepared competent cells and TCM (10 mM $CaCl_2$, 10 mM $MgCl_2$, 10mM Tris-HCl pH 7.5) were defrosted and the ligation solution was added. The mixture was incubated on ice for 30 minutes and then transferred for 3 minutes to 42ºC. After that, the cells are incubated on ice for 5 minutes. Next, 800 µL of LB medium was added and the cells are incubated with aeration for 1 hour at 37ºC and 250 rpm. The transformed *E. coli* cells were then plated in LB agar medium supplemented with 150 µg/mL Ampicillin (Sigma Aldrich, USA), a selection marker of the plasmid, and incubated overnight at 37ºC. To confirm the presence of the fragment on the plasmid and the subsequent transformation, a colony PCR (*Taq polymerase*, NZYTech, Portugal) was performed. Desired colonies were picked, resuspended in 50 µL of water, and subjected to 100ºC for 10 minutes. The mixture was then centrifuged for 1 minute at 21500 *g* and 1 µL of supernatant is used in the colony PCR with a reaction volume of 10 µL. The primers used for the *bcam0695* gene were used (Table 1), as well as the same program described before. After confirmation, the pDrive695 plasmid was extracted from one colony using the NZY MiniPrep kit (NZYTech, Portugal), following the manufacturer's instructions, and stored at -20ºC. The construct was sent for final confirmation by sequencing (GATC Biotech AG, Germany).

After completing the construction of the plasmid pDrive695, the next step was to amplify by PCR (Platinum *Taq* polymerase, Invitrogen, Thermo Fischer Scientific, USA) the 1.1-kb fragment from the pUC-Tp plasmid containing the Trimethoprim (Tp) cassette using the primers (0.2 µM) designed using the AmplifX1.5.4 software: frTp_Fwd and frTp_Rev (Stabvida, Portugal) (Table 1) possessing *XhoI* restriction sites (underlined). The PCR was performed in a reaction volume of 25µL, with $MgCl_2$ (1.5 mM), Buffer (1 x), DMSO (8% (v/v)), dNTP (0.02 mM) and pUC-Tp DNA (58 ng). The program was designed considering the enzyme manufacturer's instructions, with an annealing step at 55ºC and 35

cycles. The PCR product was confirmed with a 0.8% agarose electrophoresis gel and the Tp cassette was then extracted from the gel using the NZY GelPure kit (NZYTech, Portugal) following manufacturer instructions. The insert and the pDrive695 vector were then digested (~500ng of DNA in each digestion) using the restriction enzyme *XhoI* (Takara, Japan) in a total reaction volume of 20µL and kept for 5 hours at 37ºC. After digestion, the vector was treated with 1 unit of CIAP (Invitrogen, Thermo Fischer Scientific, USA), for 15 minutes at 50ºC and 15 minutes at 65ºC, to avoid recircularization. The Tp insert and the pDrive695 vector were purified using the Zymo Research DNA Clean and Concentrator kit (Zymo Research, USA). The pDrive695 vector and the Tp insert were ligated with 1 unit of T4 DNA Ligase (Invitrogen, Thermo Fischer Scientific, USA), vector: insert ratio of 1:3 and kept at room temperature for 30 minutes, and then transferred to a thermos for a progressive decrease in temperature and placed at 4ºC overnight. The ligation was then used to transform chemically competent *E. coli* αDH5 cells that were plated in an LB agar medium supplemented with 100 µg/mL of Ampicillin (Sigma Aldrich, USA) and 100µg/mL trimethoprim (Sigma Aldrich, USA). The transformations were kept at 37ºC overnight. To confirm the ligation and transformation, a colony PCR (*Taq polymerase*, NZYTech, Portugal) was performed using the primers designed for the BCAM0695 fragment, with the same program as described before, considering the enzyme manufacturer's instructions. The pDrive695Tp plasmid was then extracted using the NZYTech Mini Preps kit (NZYTech, Portugal), and a PCR for confirmation of the Tp insert was performed, using the *bcam0695* primers and with the same program referred previously.

The pDrive1524Tp plasmid was previously constructed and available in the lab (João Dias and Dalila Mil-Homens, unpublished). The protocol used for this construction was similar to the one used for the *bcam0695* gene.

The pDrive695Tp and pDrive1524Tp plasmids were purified using the Zymo Research DNA Clean and Concentrator kit (Zymo Research, USA) and then inserted in pre-prepared electrocompetent *B. cenocepacia* K56-2 cells by electroporation. For this, *B. cenocepacia* K56-2 cells were defrosted on ice and the DNA was introduced by electroporation. This step was repeated several times using different quantities of plasmid ranging from 60 to 1000 ng and different resistance values (200 and 400 Ω). The transformants were selected on LB agar supplemented with 150 mg/L of trimethoprim at 37ºC and different periods of incubation were tested: from 24 hours to 10 days. To distinguish between single- and double-crossover mutants, the genomic DNA was extracted using the Qiagen DNAeasy kit (Qiagen, Germany) following manufacturer's instructions, and a PCR using the respective primers (either for *bcam0695* or *bcal1524*) (Table 1) was performed, allowing the identification of the mutants.

Only the *bcal1524* mutant was achieved and it was confirmed by whole-genome sequencing using an Illumina MiSeq system at Instituto Gulbenkian de Ciência (Portugal).

2.5. Growth curves

B. cenocepacia K56-2 wild-type and Δ1524::Tp mutant were grown overnight in LB medium (for the mutant supplemented with 150 mg/L of trimethoprim) at 37ºC and 250 rpm. The cultures were diluted in LB to an OD_{600} of 0.1. The wild-type and mutant cultures were then incubated for 27h at 37ºC and 250

rpm, with the OD_{600} and viability being followed using a spectrophotometer (Hitachi U-2000, USA) and determining colony-forming units (CFU), respectively. The CFUs were determined by collecting samples, preparing serial dilutions in phosphate buffer saline (PBS) (NZYTech, Portugal), and plating 5 µL spots in LB agar plates that were then incubated for 24h at 37ºC.

2.6. Motility assays

Swimming agar plates contained 0.3% (w/v) bacto agar (Difco, USA), 10 g/L tryptone (Gibco, USA), 5 g/L NaCl (Sigma Aldrich, USA), 5 g/L yeast extract (Difco, USA). Swarming plates composition was similar, with higher bacto agar concentration – 0.5 % - and supplemented with 5 g/ L of glucose (Sigma Aldrich, USA). Bacterial cultures (wild-type and mutant) were grown overnight under defined conditions and a 5 µL drop of culture was inoculated in each plate. The plates were incubated at 37ºC for 24h and the halo diameter was determined.

2.7. Biofilm formation assay

The wild-type and mutant *B. cenocepacia* K56-2 cultures were grown in LB medium overnight at 37ºC and 250 rpm. Cultures were diluted in LB medium until an OD_{600} of 0.05 and distributed in a 96-well U-bottom polystyrene microplate (Orange Scientific, Belgium) (200 µL/well). Plates were statically incubated for 24h or 48h at 37ºC. After that, the culture medium with unattached bacterial cells was gently removed and each well was carefully washed three times with PBS (NZYTech, Portugal) (200 µL) and 1% (w/v) crystal violet solution (Sigma Aldrich, USA) was applied to stain the adherent bacteria, incubating it for 15 minutes at room temperature. After incubation, the stain was removed, and the wells were once again carefully washed three times with PBS. Ethanol (96% v/v) was used to solubilize the stain, mixing gently, and the absorbance was measured in a microplate reader (Versamax, Molecular Devices, UK) at 595 nm. The results presented correspond to mean values of at least 5 replicates from three independent experiments.

2.8. Bacterial adhesion to human bronchial epithelial cells

The adhesion experiment was performed on 16HBE14o- cells following the procedure described by Mil-Homens and colleagues[119], with some modifications. Briefly, the human bronchial epithelial cells were seeded on a 24-well plate (Orange Scientific, Belgium) at $5x10^5$ cells/well in MEM medium supplemented and incubated for 24h at 37ºC, in a humidified atmosphere with 5% CO_2. The wells were then washed with Hepes Buffer Saline (HBS) pH 7.4 (Sigma Aldrich, USA) and maintained in MEM medium, without supplementation. The wild-type and mutant strains were grown overnight in previously described conditions and were used to infect the epithelial bronchial cells at a multiplicity of infection (MOI) of 50:1 (bacteria to human cell). After infection, the plates with cells were incubated at 37ºC, in a humidified atmosphere with 5% CO_2 for 30 minutes. Next, the MEM medium was removed, the cells were washed three times with PBS and 200 µL of lysis buffer containing 0.25% (v/v) Triton X-100 (Sigma Aldrich, USA) and 10 mM EDTA (Sigma Aldrich, USA) was added and incubated for 20 minutes at room temperature. The cell lysates were scraped and mixed with the lysis buffer. For quantification, serial

dilutions of the cells' lysate mixture in PBS were performed and plated in LB agar plates. Results are presented as the percentage of adhered bacteria relative to the initial dose applied.

2.9. Adhesion of bacterial cells to ECM components

To test the bacterial ability to adhere to extracellular matrix (ECM) components, the procedure was performed as described by Mil-Homens and colleagues[59]. Briefly, three components were selected: fibronectin and collagen type I and IV. Initially, 96-well polystyrene U-bottom plates (Scientific Orange, Belgium) were coated with 10 µg/mL of the referred ECM components in PBS (NZYTech, Portugal) and placed at 4ºC overnight. The wells were washed twice with PBS and saturated with bovine serum albumin (BSA) (2% w/v) (Sigma Aldrich, USA), to block unspecific bonding, incubating for 1 hour at room temperature and being then removed. Bacterial cultures were grown in LB medium (supplemented with 150 mg/ml of trimethoprim for the mutant strain) overnight at 37ºC and 250 rpm. A bacterial suspension of $5x10^6$ CFU/mL in PBS was prepared and added to the coated wells and incubated statically for 2 hours at 37ºC. The suspension with unbounded bacteria was removed, the coated wells were washed twice with PBS and the bound bacteria were treated with 200 µL of lysis buffer containing 0.25% (v/v) Triton X-100 (Sigma Aldrich, USA) and 10 mM EDTA (Sigma Aldrich, USA), for 2 hours at room temperature, to desorb the bound bacteria. Finally, the lysates were serial diluted in PBS and plated in LB agar plates. The results are presented as the percentage of bound bacteria relative to the administered dose and are mean values of at least 5 repeats from 3 independent experiments.

2.10. *Galleria mellonella* killing assays

Galleria mellonella wax moth larvae were reared in our lab at 25 ◦C in the dark, from egg to last-instar larvae, on a natural diet (beeswax and pollen grains). Worms of the final-instar larval stage, weighing 225 ± 25 mg, were selected for the experiments. *G. mellonella* killing assays were performed based on a previously described method, with some changes[129]. Cultures of *B. cenocepacia* K56-2 and the Δ1524::Tp mutant strain were grown overnight at 37ºC, 250 rpm. Cultures were centrifuged (6120 g, 2 minutes), the supernatant discarded, and the cells resuspended in PBS. Serial 10-fold dilutions in PBS were performed until obtaining a concentration of $2x10^3$ CFU/mL. Using a micrometre adapted to control the injection volume of a micro-syringe, 5 µL of bacteria suspension was injected into *G. melonella* via the hindmost left proleg, containing 10 CFU/injection. Control larvae were injected with PBS and 5 µL of bacterial suspension was plated in LB agar to control the number of CFU injected. After injection, the larvae were placed in Petri dishes and stored in the dark at 37ºC. The survival was recorded at 24 hours intervals until the 72- hour time point. Caterpillars were considered dead when they displayed no movement in response to touch.

2.11. Production and purification of recombinant truncated proteins

The initial strategy to produce our recombinant proteins was to use the pET28-a (Novagen, Sigma-Aldrich, USA) plasmid as an expression vector in *E. coli* BL21-DE3. For that, the *bcam0695* gene (encoding for 737 amino acids, excluding 21 aa from the lipid box identified with the software Cleavage LipoP1.0) was amplified using *B. cenocepacia* K56-2 genomic DNA (45 ng) as a template and primers

frBCAM0695_21_758_fwd, and frBCAM0695_21_758_rev (Stabvida, Portugal) (Table 1), containing, respectively, *NdeI* and *BamHI* restriction sites (underlined). The PCR was performed using *Platinum Taq* DNA polymerase (Invitrogen, Thermo Fisher Scientific, USA), with $MgCl_2$ (1.5 mM), Buffer (1x), dNTP (0.2 mM), both primers (0.2 μM) and DMSO (8% (v/v)). After that, the PCR product was confirmed with a 0.8 % agarose electrophoresis gel and the obtained fragment was directly purified from the PCR reaction mixture using the Zymo Research DNA Clean and Concentrator kit (Zymo Research, USA), following the manufacturer's instructions. Then, it was digested (~1000 ng of DNA) with *NdeI* (Takara, Japan) and *BamHI* (Takara, Japan) restriction enzymes in a reaction volume of 20μL. At the same time, the pET-28a (Novagen, Sigma-Aldrich, USA) plasmid was digested with the same restriction enzymes also in 20 μL of reaction volume. Both digestions were kept at 37ºC overnight. The pET-28a plasmid was treated with CIAP (Invitrogen, Thermo Fisher Scientific, USA) for 10 minutes at 37ºC and 15 minutes at 65 ºC, to avoid recircularization. Both digestions were purified using the NZY GelPure kit (NZYTech, Portugal) and afterwards, the BCAM0695_21_758 fragment was cloned into the pET-28a plasmid with a 6xHis tag in its N-terminus and transformed into *E. coli* αDH5 posteriorly plated in LB agar supplemented with kanamycin (50 mg/L) and incubate at 37ºC for 24h. Growing colonies were picked to confirm the presence of the construction by colony PCR as previously described. The selected colony was then used to extract the construction using the NZYMiniprep kit (NZYTech, Portugal). The construct was then inserted into *E. coli* BL21-DE3, by electroporation. To confirm the correct transformation, cells were plated in LB agar supplemented with kanamycin (50 mg/L). To overexpress the BCAM0695_21_758 fragment, cells were incubated in LB medium overnight at 37ºC with 50 mg/L of kanamycin. The culture was then diluted until reaching an OD_{640} of 0.1 in LB medium with kanamycin. The cells were grown until an OD_{640} of 0.6 and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (NZYTech, Portugal) at 0.2, 0.6 and 1 mM for 4 hours at 37ºC, 250 rpm. The same procedure was performed to test other expression conditions: after induction with IPTG, the cells were incubated overnight at 16ºC, 200 rpm; after the overnight growth, cells were transferred to SB medium (32 g/L tryptone, 20 g/L yeast extract, and 5 g/L NaCl), induced with IPTG and incubated for 4h at 37ºC.

In addition to the procedure executed with the plasmid pET-28a, a parallel protocol using the Champion™ pET SUMO Protein Expression System kit (Invitrogen, Thermo Fisher, USA) was performed to amplify the same BCAM0695_21_758 fragment, but also other truncates: BCAM0695_320_758, BCAL1524_33_558 and BCAL1524_116_558, derived from the *bcam0695* and *bcal1524* genes, respectively. For that, the primers for each truncate present in Table 1 were used, as well as *Platinum Taq* DNA polymerase (Invitrogen, Thermo Fisher Scientific, USA), with $MgCl_2$ (1.5 mM), Buffer (1x), dNTP (0.2 mM), and DMSO (8% (v/v)). Additionally, a different plasmid vector was used for the constructions: pET SUMO, from the Champion™ pET SUMO Protein Expression System (Invitrogen, Thermo Fisher, USA). The construction of the vectors was performed with the fresh PCR products and the insertion into chemically competent *E. coli* αDH5 following manufacturers' instructions. The correct construction was confirmed by sequencing (GATC Biotech AG, Germany). Posterior transformation of chemically competent *E. coli* BL21-DE3 cells was performed. The truncates from the *bcam1524* gene were overexpressed testing the following conditions: induction with IPTG (0.2, 0.6 and 1 mM) for 4 hours at 37ºC, 250 rpm. For each time point and concentration of IPTG a sample of cells

was harvested and visualized in a 10% SDS-Page gel. For the optimal expression conditions, cells were harvested from 1L of culture by centrifugation at 7025 g for 10 minutes at 4ºC. The supernatant was discarded, and the pellet was resuspended in buffer A (20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 7.4) and the cells were disrupted by sonication using the Branson Sonifier 250 (Emerson, USA) (8 cycles of 15 pulses with 70% duty cycle and output control 8) and centrifuged at 17600 g for 5 minutes and then the supernatant was transferred to new tubes and centrifuged again for 1 hour. The supernatant was collected and loaded into a 5 mL HisTrap column (GE Healthcare, Germany) equilibrated with buffer A using an AKTA Start system (GE Healthcare, Germany), following the manufacturer's instructions. The amount of imidazole was altered by a continuous gradient from 10 mM to 300 mM. After elution, samples collected were analysed by SDS-Page to confirm the fractions with our fragments of interest (BCAL1524_33_558 and BCAL1524_116_558). Those fractions were then concentrated using an Amicon 10kDa (Merck Millipore, Ireland) and dialysed to PBS. The concentrated samples were quantified using the Nanodrop One (Thermo Fisher, USA) and stored at 4ºC. The next step was to separate the SUMO protein from our fragments of interest. For that, the SUMO protease was added (0.1% v/v) and the mixture was incubated overnight at 4ºC. Then, the mixture was loaded into a 1 mL HisTrap column (GE Healthcare, Germany) to collect the flowthrough that contain our fragments, using buffer A. Then SUMO protein was eluted with a phosphate buffer containing 300 mM imidazole. Both fractions, one containing the fragments and the other containing SUMO protein were analysed by SDS-Page to confirm the correct separation. The fractions with the fragments were concentrated using an Amicon 10kDa (Merck Millipore, Ireland), dialysed to PBS, quantified using the Nanodrop One (Thermo Fisher, USA) and stored at 4ºC.

2.12. Statistical Analysis

All experiments were performed in a minimum of three independent replicates. Statistical analysis was carried out by using GraphPad Prism 8-0-1 software. Relative comparisons were done among corrected values with the Welch's t test for significance. For the animal model, Kaplan-Meier survival curves were plotted, and differences in survival rates were analyzed using a log-rank (Mantel-Cox) statistical test. A p-value of < 0.05 was considered statistically significant in all analyses.

Table 1. List of primers used in this study.

frBCAM0695_fwd	5'-CGGGATCCACAACCTCGGCAATGCAGTGA-3'
frBCAM0695_rev	5'-TACGTCTAGAGTCGTTCTGGTACGTGATGCGTTA-3'
frTp_fwd	5'-ACCGCTCGAGCGCCCACAGTCCATTGAACAAA-3'
frTp_rev	5'-ACCGCTCGAGTATGCTTCCGGCTCGTATGTTG-3'
frBCAL1524_fwd	5'- CGGGATCCAACCGTTCGCCACATCTGCT- 3'
frBCAL1524_rev	5' - TACGTCTAGATGATCGCCGAGCTGATCGGTT- 3'
frBCAM0695_21_758_fwd	5'-GGGAATTCCATATGTACGGTTGCGGCTCGGTCGAT-3'
frBCAM0695_21_758_rev	5'- CGCGGATCCTGTCGTTCTGGTACGTGATGCGTT-3'
frBCAM0695_21_758_fwd2	5'-TACGGTTGCGGCTCGGTCGAT-3'
frBCAM0695_21_758_rev2	5'-TGTCGTTCTGGTACGTGATGCGTT-3'

frBCAM0695_320_758_fwd	5’-CAGAAATCCGGCAACCTCGTGA-3’
frBCAL1524_33_558_fwd	5’- GGCTCCATCAGCCAGGGTCT-3’
frBCAL1524_116_558_fwd	5’-GTGGGAAATGTCCTCGCACAA-3’
frBCAL1524_33_558_rev	5’-CCTCGCTGTCTTTTCGATGAATGG-3’

3. Results and Discussion

3.1. Bioinformatic analyses of collagen-like proteins in the *Burkholderia cepacia complex*

Collagen-like proteins are being identified in increasing numbers, namely in pathogens[68,71,77]. In this work, we began by identifying the presence and distribution of Collagen-like proteins among a representative strain for each of the 19 members of the Bcc using the Burkholderia Genome Database. For reference, the Bacterial Collagen Middle Region (*Col_mid_reg*) (PF15984 in InterPro) was used. This domain is described as a conserved domain represented in bacterial collagen, and is mainly present in bacterial species, as represented in Figure 4 A. However, there are 8 species in the Eukarya domain. Inside the Bacteria domain, the Betaproteobacteria class is the one with the most species, being registered 57 species out of the 85 registered in the Bacteria domain. Finally, the Burkholderiaceae family, which englobes the *Burkholderia* genus and the Bcc, is the family with the most species represented, with a total of 38. This indicates the existence of a higher number of proteins with the *Col_mid_reg* domain, which may also indicate the existence of more collagen-like proteins yet to study. Considering the search for collagen-like proteins in Bcc, a total of 75 putative collagen-like proteins were identified, whose sequences were then used to obtain the phylogenetic tree represented in Figure 4 B. The phylogenetic tree is divided into four different clusters (I, II, III and IV), with cluster IV subdivided into three subclusters (IV_A, IV_B, and IV_C). Analysing the tree, it is understandable that the sequences are not grouped by species, but the division is rather based on the identity of the *Col_mid_reg*. Thus, proteins from the same cluster possess identities of around 80-90%, while proteins from different clusters can have identities of 40%. It is notable that sequences from the same species are distributed through different clusters, which indicates that the *Col_mid_reg* is variable within the same species. This variability may indicate that this domain may perform different roles, with different types of *Col_mid_reg* having different affinities to different receptors. However, more studies are necessary to fully understand the importance of this domain.

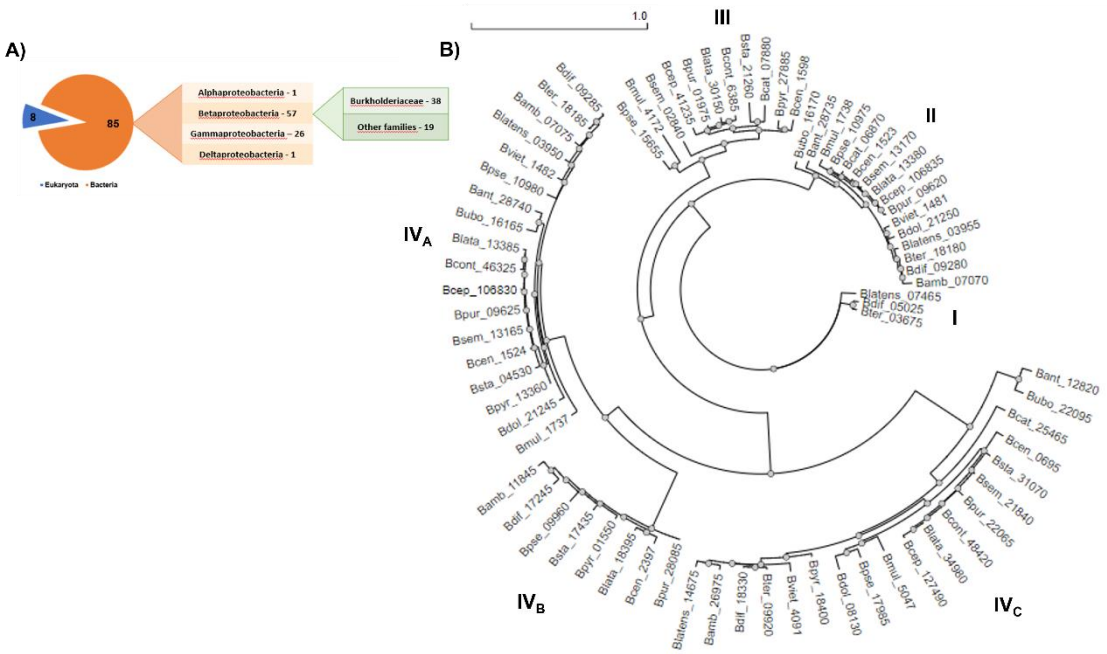


Figure 4. A) Distribution of the Bacterial Collagen Middle Region through different levels of taxonomy: domains, classes, and families, respectively. **B)** Phylogenetic tree of the 75 putative collagen-like proteins from the Bcc, divided into four clusters: I, II, III and IV. The division is based on the identity of the *Col_mid_reg*.

Next, an analysis regarding the presence and frequency of Tandem Repeats (TR) in these 75 proteins was performed, due to the already described importance of tandem repeats in adhesion and infection processes[130,131]. Using the XSTREAM web server, the tandem repeats were identified, and the total percentage of the protein occupied by these repetitions was calculated. This is represented in Figure 5. Firstly, it is notable the existence of tandem repeats in the majority of the 75 proteins, being only absent in 5 proteins (6.67 %). Additionally, in this graph is visible that *B. cenocepacia* J2315 and *B. multivorans* ATCC 17616 possess all the proteins with tandem repeats. These two species are known for being the two most prevalent Bcc pathogenic species[132]. *B. cepacia* ATCC 25416 is also pathogenic[133] but possesses a low percentage of tandem repeats in comparison with the other proteins analysed, being for that reason difficult to associate the tandem repeats to an increase in pathogenicity in these species without further studies, even though there seems to be a tendency in that sense.

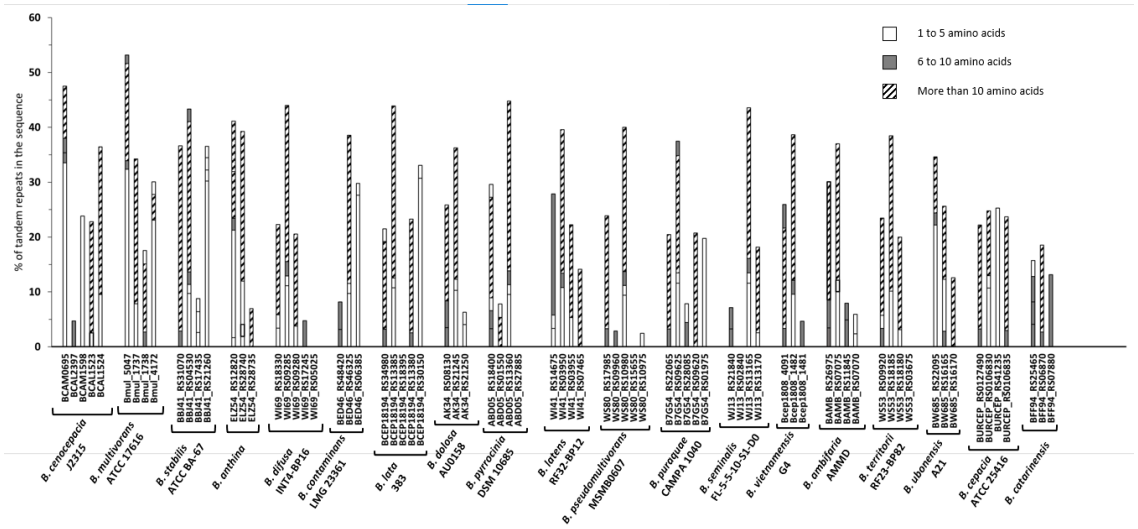


Figure 5. Distribution and frequency of tandem repeats in the 75 collagen-like proteins from the Bcc.

We next analyzed the sequences' distribution by species and by subcellular localization of the product (Figure 6). We can observe that the number of CLPs per species is between 3 and 5 and that the majority are predicted as membrane proteins. More specifically, there are a considerable number of proteins with predicted localization in the cytoplasmic membrane or the outer membrane. Is it also observable that there is at least one protein possessing a lipoprotein domain, except in *B. vietnamensis* and *B. dolosa*. This even distribution of collagen-like proteins in all the members of the Bcc may indicate the importance of these proteins in indispensable biological functions for the species, namely to adhesion, due to the prevalence of these proteins in the membrane.

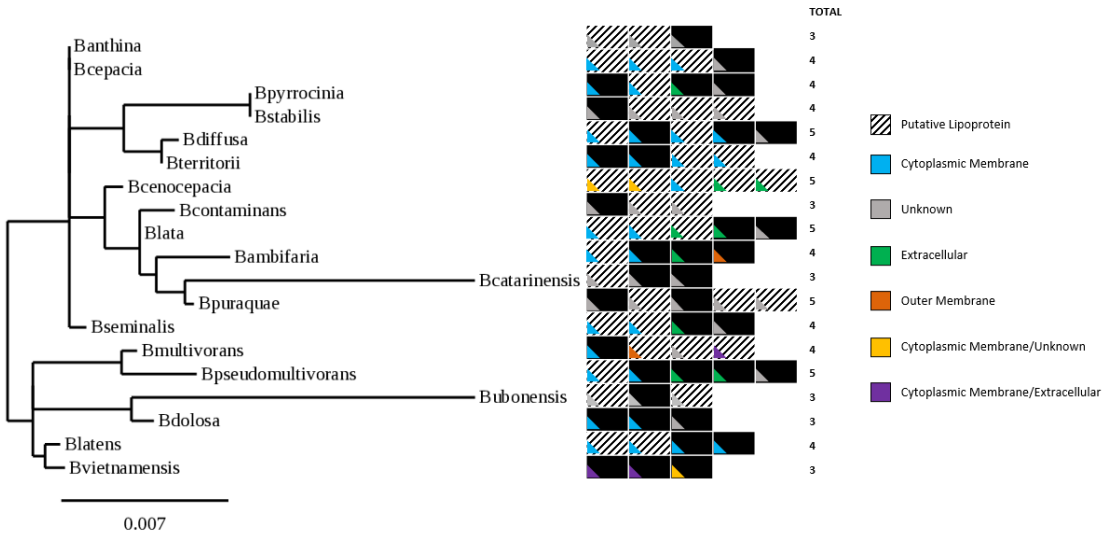


Figure 6. Distribution of the 75 collagen-like proteins per species and in subcellular localization.

Next, we focused our analysis on *B. cenocepacia* J2315 (ET-12 lineage, it is considered the reference strain), 5 collagen-like proteins were identified: BCAL1523, BCAL1524, BCAL2397, BCAM0695, and BCAM1598. These proteins are all characterized as putative lipoproteins, indicating their potential as having a role in the adhesion process. Additionally, their cellular localization is described as in the cytoplasmic or extracellular membrane, as expected due to their lipoprotein nature, and the genes

bcal1523, bcal1524 and *bcam0695* appeared overexpressed in Pimenta et al study[29], as already referred, predicting their importance in the adhesion process. We then focused on these 5 collagen-like proteins and genes from *B. cenocepacia* J2315.

Firstly, we assessed the distribution of the sequences on the chromosomes. *B. cenocepacia* J2315 possesses 3 chromosomes and 1 plasmid[132], although the 5 genes of interest are distributed only in chromosome 1 – the "largest" - (*bcal1523, bcal1524,* and *bcal2397*) and chromosome 2 – the "medium" – (*bcam0695* and *bcam1598*) (Figure 7 A). All were found on the minus strand, except for *bcal2397*, found on the plus strand. The next step was to analyze the genes' neighbourhood, consulting the Burkholderia Genome Database (Figure 7 B and Table 2. Description of the CLPs neighborhood genes, with gene name, genomic location, strand, product description and cellular localization.). Most of the genes rightly up and downstream our genes of interest are not characterized yet, possessing any function attributed. However, two genes (*bcal1520* and *bcal1525*) near *bcal1523* and *bcal1524* were also overexpressed in Pimenta's experiment[29] which may indicate their importance and coordination with our target collagen-like proteins. Noticeably, the *bcal1520* product is also described as a putative lipoprotein and *bcal1525* is already described as an Flp-type Pilus subunit (Table 2), being a virulence trait described in other organisms[134,135]. Studying the up and downstream genes around the collagen-like encoding genes is important to assess the eventual coordination and existence between them, also due to their proximity in the genome.

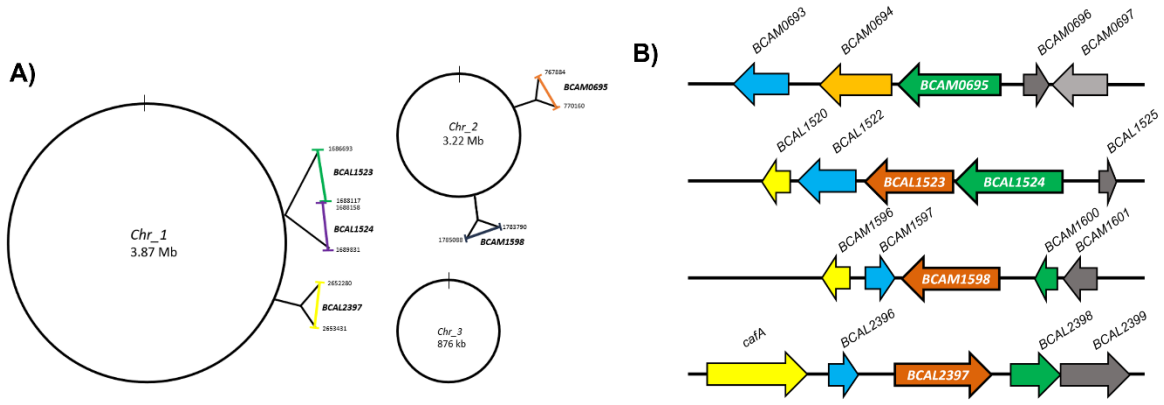


Figure 7. A) Distribution and localization of the 5 collagen-like encoding genes in the *B. cenocepacia* J2315 chromosomes. **B)** Neighbourhood of the 5 collagen-like proteins.

Table 2. Description of the CLPs neighborhood genes, with gene name, genomic location, strand, product description and cellular localization.

Gene Name	Genomic Location	Strand	Product Description	Cellular Localization
BCAM0693	765308-765925	-	Cytoplasmic	Hypothetical Protein
BCAM0694	766473-767756	-	Cytoplasmic Membrane	Hypothetical Protein
BCAM0695	767884-770160	-	Cytoplasmic Membrane/Unknown	Putative Lipoprotein
BCAM0696	770523-770921	+	Unknown	Putative Carboxymuconolactone Decarboxylase
BCAM0697	770994-771977	-	Unknown	Putative Hydrolase
BCAL1520	1684227-1684688	-	Periplasmic	Putative Lipoprotein
BCAL1522	1684863-1686548	-	Outer Membrane	Putative Exported Heme Utilisation Related Protein
BCAL1523	1686693-1688117	-	Cytoplasmic Membrane	Putative Lipoprotein
BCAL1524	1688158-1689831	-	Extracellular	Putative Lipoprotein
BCAL1525	1690509-1690679	+	Unknown	Flp type Pilus Subunit
BCAM1596	1782467-1783099	-	Cytoplasmic Membrane	LysE Family Transporter
BCAM1597	1783216-1783677	+	Cytoplasmic	AsnC Family Regulatory Family
BCAM1598	1783790-1785088	-	Cytoplasmic Membrane/Extracellular	Putative Lipoprotein
BCAM1600	1785332-1785679	-	Periplasmic	Hypothetical Protein
BCAM1601	1785732-1786523	-	Cytoplasmic Membrane	Hypothetical Protein
cofA	2649893-2651362	+	Cytoplasmic	Ribonuclease G
BCAL2396	2651489-2651824	+	Unknown	Hypothetical Protein
BCAL2397	2652280-2653431	+	Extracellular	Putative Lipoprotein
BCAL2398	2653685-2654650	+	Unknown	Putative Lipoprotein
BCAL2399	2654647-2655906	+	Cytoplasmic Membrane	Major Facilitator Superfamily Protein

Afterwards, the characterization of the five CLPs was performed. Using InterPro and the SMART web servers, we identified the size and localization of the Col_mid_reg and of low complexity regions along the protein, respectively (Figure 8). The SMART webserver also identified the existence of a transmembrane domain for BCAL2397. The Bacterial Collagen Middle Region is present in all 5 proteins, as expected, with a size of approximately 200 amino acids, and all 5 proteins possess several low complexity regions, especially in the areas not identified as the Col_mid_reg (Figure 8). In fact, the percentage of protein considered with low complexity is variable, with a maximum of 72 % for the BCAM0695 protein and a minimum of 43 % for the BCAL1523 (Table 3). Moreover, we searched for the presence of GXY repeats, typical of collagen-like domains and responsible for the formation of triple-helix structures. Manual visualization allowed the identification of GXY repeats in 4 out of 5 proteins, with BCAL2397 as the exception. The number of repeats varies between four in BCAL1523 and 83 in BCAM0695. Considering the genes overexpressed in Pimenta et al. study[29], we can see that BCAL1524 possesses 18 GXY repeats, besides the mentioned 83 repeats for BCAM0695 (Table 3). BCAM1598 possesses 36 repeats. A computational search of the GXY repeats was performed, but only domains with 20 copies of the GXY triplet are detected by InterPro, indicating that proteins with fewer may go unnoticed. Thus, only BCAM1598 and BCAM0695 possess enough repetitions to be detected by Interpro. Analyzing the sequences from the four G-X-Y containing proteins, we can observe that the most frequent combination of amino acids is G-T-S (Glycine – Threonine – Serine), comprising 91% of

the G-X-Y tripeptides present in the proteins. The number of repetitions and the most typical pattern is by what was described by Qiu and colleagues, as the maximum number of G-X-Y repeats observed in organisms from the Proteobacteria phylum was 406 and the most common amino acid composition of the triplet was registered as G-T-S[82]. It is also noted that the GXY domains identified by InterPro and manually confirmed are located upstream of the collagen middle region present in all the CLPs (not shown) and that no other domains were detected by the software. Focusing on the location of the GXY repeats and the existence of low complexity regions along the proteins, we can observe that coincidence. This may be due to the nature of the amino acids and of the repetitiveness of that area of the protein. It is known that the presence of a higher frequency of Glycine and Serine, among others, can lead to the formation of disordered regions in proteins, something that is also related to collagen domains[86]. Additionally, it is notable that the protein with the highest percentage of low complexity regions (BCAM0695) is also the protein with the highest number of GXY repeats, indicating the correlation between these two characteristics, even though the presence of low complexity regions is not completely dependent of the presence of GXY, as we may see in proteins BCAL1523 and BCAL2397, which possess few or no repetitions and present low complexity regions throughout the extension of the sequence.

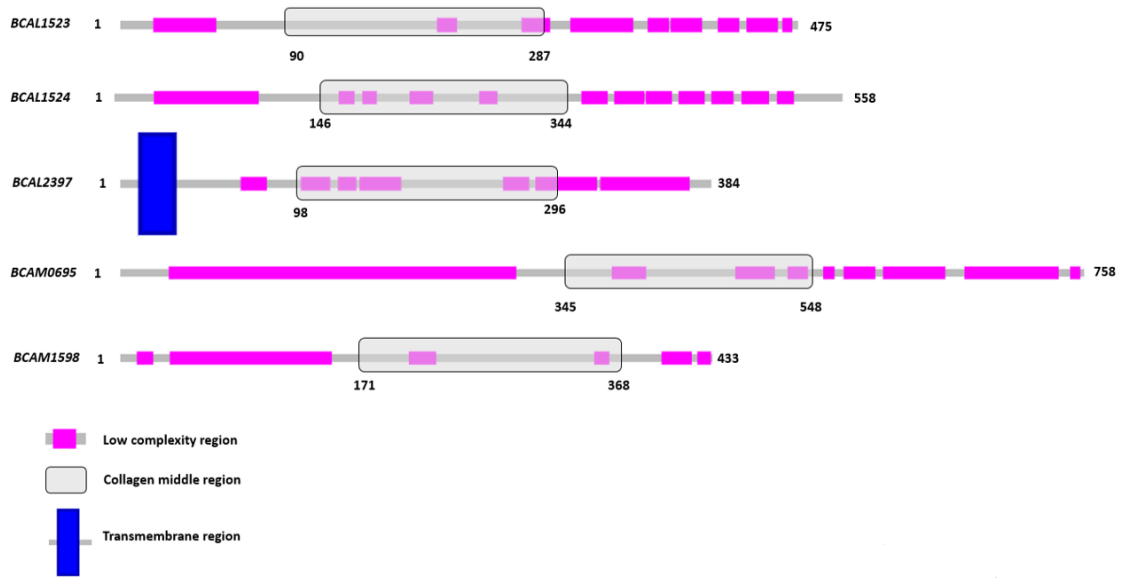
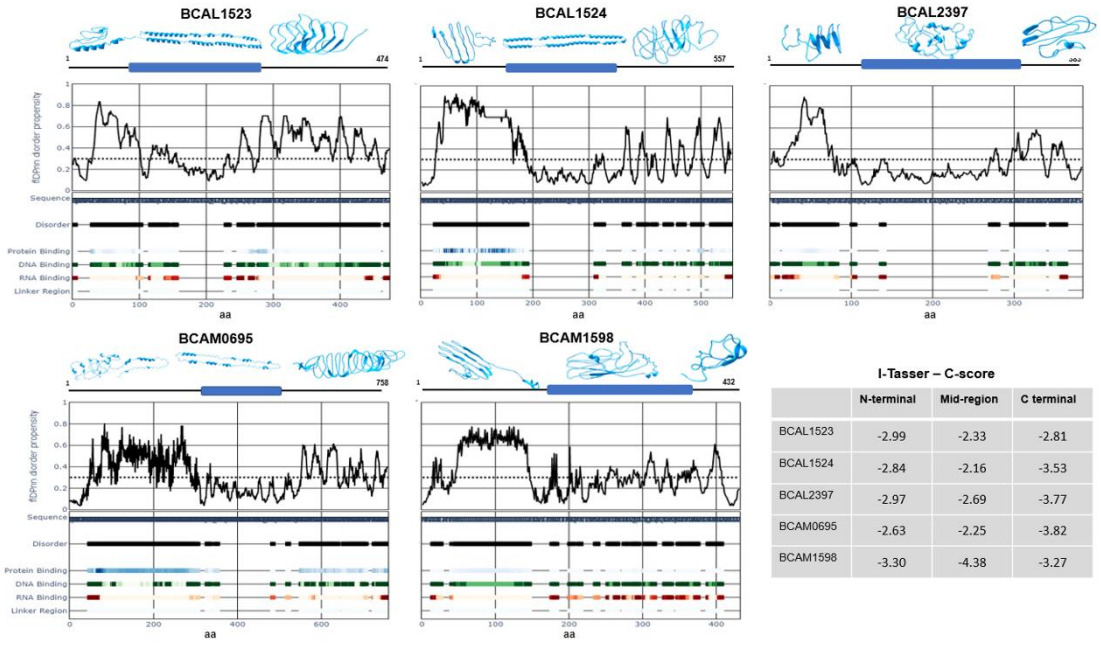


Figure 8. Representation of the low complexity regions and the bacterial collagen middle region predicted by SMART and InterPro webservers, respectively.

Table 3. Characterization of the 5 CLPs identified in *B. cenocepacia* J2315. The number of amino acids, the predicted localization of the protein, the significance of the existence of the Collagen Middle Region, the percentage of low complexity region per protein, and the number and type of GXY repeats.

Protein	No. amino acids	Localization	Collagen Middle Region E-value	Low complexity region (%)	No. GXY repeats	GXY Type
BCAL1523	475	Cytoplasmic Membrane	$1.8e^{-63}$	43	4	GTS_3GSS_1
BCAL1524	558	Extracellular	$5.1e^{-68}$	49	18	$GSS_1GTS_{16}GVS_1$
BCAL2397	384	Extracellular	$9.4e^{-58}$	63	-	-
BCAM0695	758	Unknown/Cytoplasmic Membrane	$9.3e^{-56}$	72	83	$GSS_3GTS_{79}GTG_1$
BCAM1598	433	Extracellular/Cytoplasmic Membrane	$1.2e^{-56}$	45	36	$GTN_1GTG_1GTS_{31}GTP_1$ GTN_1GII_1

The existence of an elevated percentage of low complexity regions pushed us to an analysis regarding the disorder of the proteins. Using the flDPnn webserver we searched for the disordered regions of the 5 CLPs and observed that the lowest disorder score coincided with the location of the collagen middle region (Figure 9). Moreover, the remaining sequence of the protein, both in the N-terminal and the C-terminal, may be considered disordered regions. The existence of a higher frequency of tandem repeats (Figure 5) in these proteins explains the presence of low complexity and disordered regions. Furthermore, the disordered regions possess a high score for protein, DNA or RNA binding, indicating their putative importance in biological functions, as described in bacteria[91]. Using the I-TASSER web server we obtained a model for each protein, which allowed the visualization of the 3 regions of each protein (N-terminal, Collagen Middle Region, and C-terminal) (Figure 9). The collagen middle regions of BCAL1523, BCAL1524 and BCAM0695 show a more defined structure, with the N- terminal and C-terminal regions possessing no defined structure, being coincident with the information obtained regarding the disordered and low complexity regions. Considering BCAL2397 and BCAM1598, even the collagen middle region possesses no defined structure. The disorder score for BCAL2397 is not high, while in BCAM1598 the disorder score may justify the representation obtained with I-TASSER. However, the quality of the predicted structures is not elevated, as expected due to the high amount of disordered and low complexity regions in these proteins (C-score table in Figure 9).



I-Tasser – C-score

	N-terminal	Mid-region	C terminal
BCAL1523	-2.99	-2.33	-2.81
BCAL1524	-2.84	-2.16	-3.53
BCAL2397	-2.97	-2.69	-3.77
BCAM0695	-2.63	-2.25	-3.82
BCAM1598	-3.30	-4.38	-3.27

Figure 9. Disorder scores obtained with the flDPnn webserver and I-TASSER structure predictions. C-score is a confidence score for the quality of the predicted structure, ranging from [-5, 2].

The presence of GXY repeats was then analyzed in the orthologues of the *bcam0695* gene, since it is known that this type of tandem repeats is prone to variation, contraction, and extensions of this area of the proteins. Looking to the genomic sequence, we observed that the tandem repeats can be classified as imperfect tandem repeats, being not totally conserved[130], even though originating conserved amino acid repetitions, due to the degenerate nature of the genetic code. It is also described that varying the extension of tandem repeats is important for the pathogenicity of bacteria, leading to a more efficient adaptation[130]. In the different members of the Bcc we observed that only the *Bmul_5047* gene from *B. multivorans* ATCC 17616 possesses the GXY repetitions conserved in the gene sequence (Figure 10). In fact, this gene possesses more nucleotide repetitions that originate GTS than *bcam0695* from *B. cenocepacia* J2315, demonstrating the existence of expansions in the type of region. However, when compared with the other members of the Bcc, no more consecutive GTS expressing repetitions are found in the gene sequence. *B. anthina* and *B. ubonensis* possess a sequence which encodes for a GTSX, with a fourth amino acid that impairs the existence of consecutive GTS repetitions. This is also common in regions enriched in tandem repeats, due to the enhanced probability of occurring mutations that alter the genomic sequence[130]. In the other members, no GTS encoding tandem repeats are annotated. This happens due to other typical phenomena in these tandem repeats: mutations that originate a frameshift. Searching upstream of the gene, we find other genes annotated possessing the tandem repeats belonging to the orthologue of *bcam0695* or the tandem repeats lost in intergenic regions. This happens due to the sequence flexibility that tandem repeats confer to the genomic sequences, especially in pathogenic organisms that take advantage of these characteristics to adapt to a hostile environment more rapidly[130]. The study of these proteins and these disordered and flexible regions is important to understand the virulence and adaptative capacity of the Bcc members, namely *B. cenocepacia* as one of the most virulent and pathogenic.

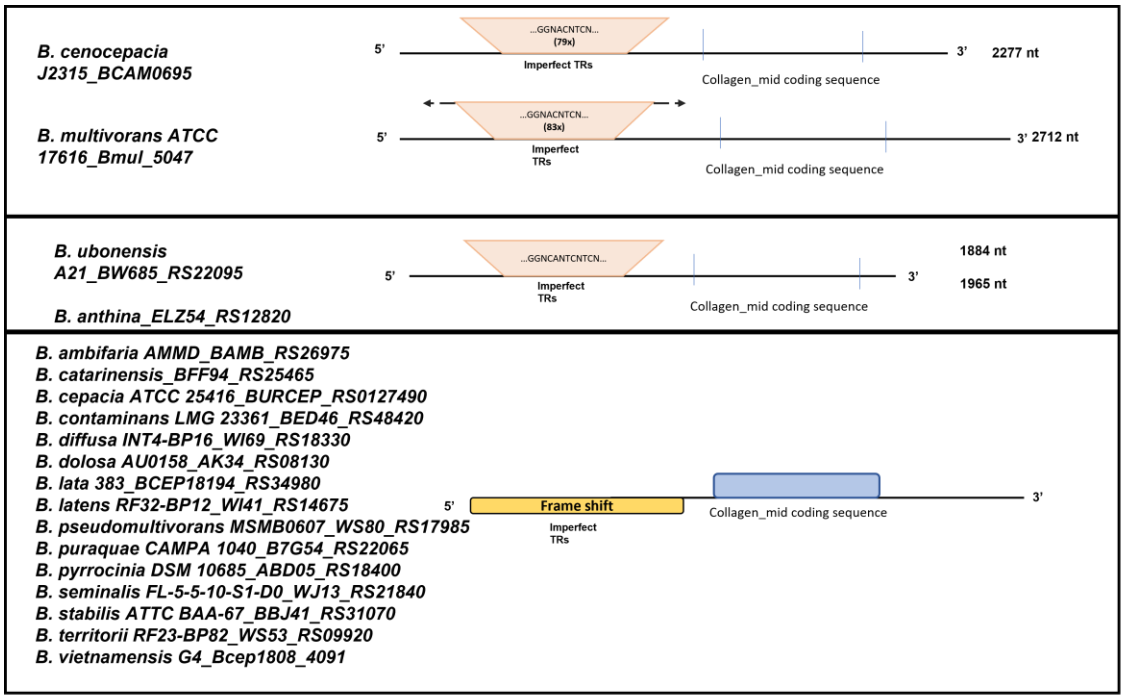


Figure 10. Distribution of the orthologues of *bcam0695* considering the existence of GTS encoding tandem repeats. In the first box, the two species possessing consecutive imperfect tandem repeats (TRs) leading to GTS. In the second box, two species with mutations leading to the loss of the GTS consecutive repetitions. In the third box, the species possessing genes with the TRs out of frame, possibly due to a frame shift mutation.

3.2. Construction of *bcam0695* and *bcal1524 B. cenocepacia* mutants

As mentioned previously, *bcal1524* and *bcam0695* are two collagen-like protein-encoding genes that had their expression augmented in the transcriptomic analysis performed after the adhesion assay performed by Pimenta and colleagues[29] against giant plasma vesicles derived from live bronchial epithelial cells, being, therefore, the two targets for the experimental section of this project. After the computational analysis, the next step was to construct a mutant for each gene individually, using the *B. cenocepacia* K56-2 strain (ET-12 lineage strain) (Figure 11).

For the *bcal1524* gene, several attempts were needed, where the concentration of plasmid was altered, ranging from 60 to 1000 ng, the resistance for the electroporation varied between 200 and 400 Ω, and the days of incubation were extended until 10 days. In one of the tries, using 250ng of plasmid pDrive1524Tp, 400 Ω of resistance, and 7 days of incubation in LB agar plate supplemented with 150 mg/L trimethoprim, a colony was observed. Next, genomic DNA was extracted and a PCR for confirmation of the insertion of the Tp cassette allowed us to confirm the construction of a Δ1524::Tp mutant in *B. cenocepacia* K56-2.

For the *bcam0695* gene, several attempts were necessary to construct the pDrive695Tp plasmid. Following, the aim was to insert the plasmid pDrive695Tp into *B. cenocepacia* K56-2 by electroporation. However, we were not able to obtain a mutant for this construction. Different concentrations and quantities of plasmid were used during the process (from 66 ng/μL to 660 ng/μL); resistance was alternated between 200 and 400 Ω; concentration of trimethoprim as a selective marker ranged from

150 mg/L to 300 mg/l; time of incubation after plated was extended until 7 days. None of these alterations allowed us to obtain a *bcam0695*::Tp mutant.

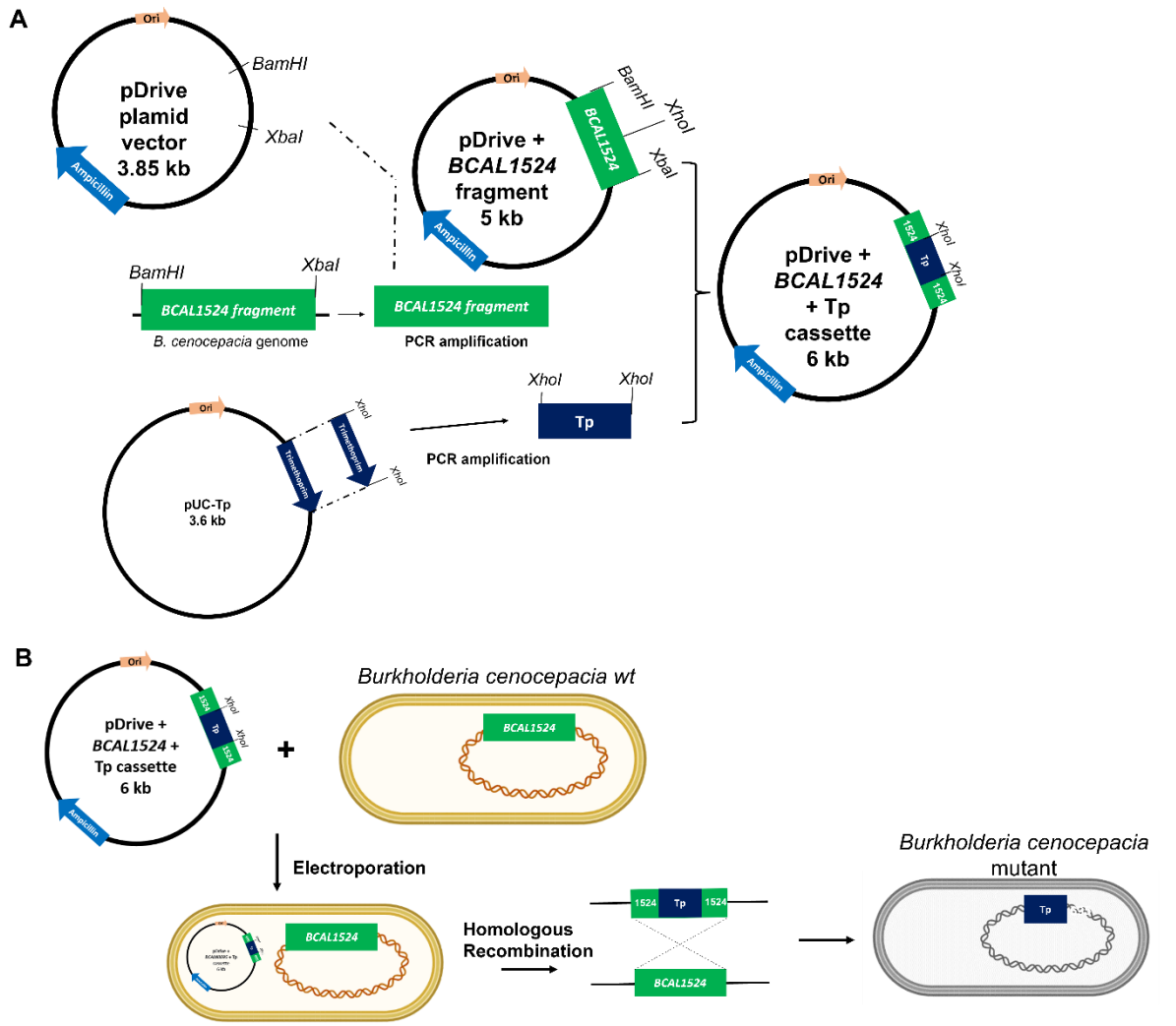


Figure 11. Process for mutant construction. **A)** The gene of interest is amplified and inserted into the pDrive vector. Then, the Tp resistance cassete is amplified and inserted in the pDrive1524 construction, interrupting the gene of interest, and creating the pDrive1524Tp construct. **B)** The pDrive1524Tp construct is inserted in *B. cenocepacia* K56-2 by electroporation and by homologous recombination the deletion mutant with the inserted Tp cassette is created.

With the Δ1524::Tp mutant, we were able to perform a phenotypic characterization compared with the wild-type strain, conducting various assays to assess the importance of this gene in growth, motility, biofilm formation, virulence, and adhesion capacity.

3.3. Effect of *bcal1524* mutation on *B. cenocepacia* K56-2 growth

The objective after constructing the mutant was to perform a phenotypic analysis of the mutant and understand the role of the BCAL1524 collagen-like protein in different processes of the bacteria, namely processes related to pathogenicity and adhesion. Thus, to assess the role of the *bcal1524* gene in the growth of *B. cenocepacia* K56-2, we compared the growth of both the wild-type and the mutant strains in LB medium by measuring the OD_{600nm} and by assessing viability through CFUs determination. As

represented in Figure 12, no significant difference in growth can be observed, which may indicate the unimportance of our gene of interest for the proper growth of this species in the tested conditions.

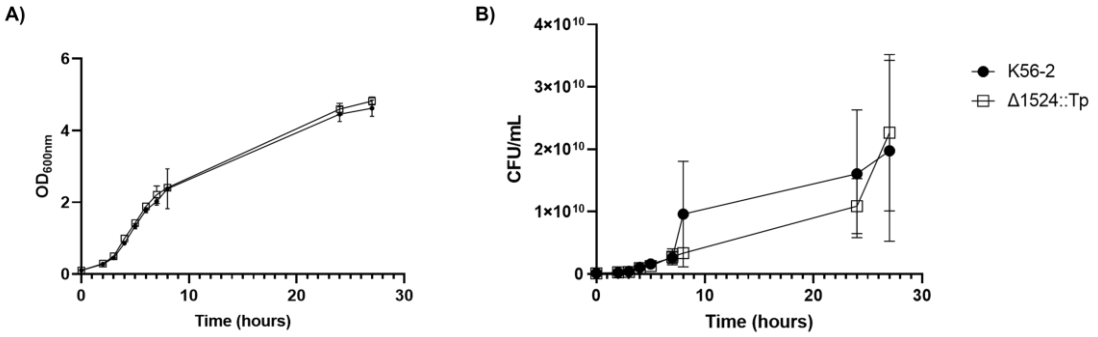


Figure 12. **(A)** Growth of *B. cenocepacia* K56-2 (•) and *B. cenocepacia* Δ1524::Tp mutant (□) in LB medium for 27 hours measured as OD_{600nm}. **(B)** Growth of *B. cenocepacia* K56-2 (•) and *B. cenocepacia* Δ1524::Tp mutant (□) in LB medium for 27 hours measured as CFU/mL. Mean values are represented with error bars corresponding to the standard deviation of 2 independent replicates.

3.4. Motility in the mutant is impaired by the *bcal1524* deletion

The motility of the bacteria was described as an important step for the initial phase of adhesion, with several studies showing that mutant bacteria with impaired motility present a decrease in adhesion, especially regarding flagellum impairment[41,136]. Collagen-like proteins were described as part of flagella in *Bacillus amyloliquefaciens* FZB42 and their absence was described as affecting motility, namely the swimming ability of the bacteria[137]. We, therefore, proceeded to analyse the swimming and swarming ability of the mutant compared with the wild-type strain. The results presented in Figure 13 A) show a significant decrease ($P < 0.01$) in the swimming halo diameter of the mutant when compared to the wild type, indicating that the BCAL1524 protein may play an important role in the motility of the bacteria. The swarming ability (Figure 13 B) is also affected, which reinforces that these results may indicate the BCAL1524 role related to motility and to flagella, although more tests are necessary to fully understand this process.

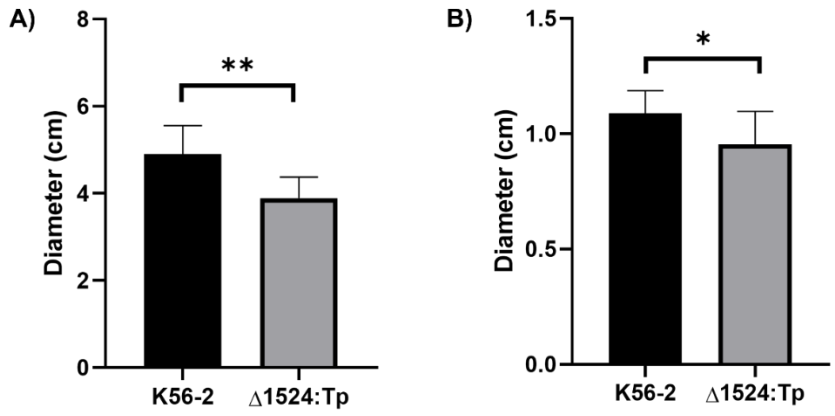


Figure 13. Motility assays. **A)** Swimming motility assay after 24h of incubation of wild-type *B. cenocepacia* K56-2 and *B. cenocepacia* Δ1524::Tp mutant. The graph translates the mean diameter with error bars corresponding to the standard deviation of the halo from three independent replicates. Swimming ability is impaired in the mutant (**, $P < 0.01$). **B)** Swarming motility assay after 24h of incubation of wild-type *B. cenocepacia* K56-2 and *B. cenocepacia* Δ1524::Tp mutant. The graph translates the mean diameter with error bars corresponding to the standard deviation of the halo from three independent replicates. Swarming ability is also altered in the mutant (*, $P < 0.05$).

3.5. Biofilm formation is not affected in the Δ1524::Tp mutant

The next step was to evaluate the biofilm formation ability of the mutant. Biofilm formation is an important trait in pathogenic organisms since it protects the bacteria from immune response or other antimicrobial agents. In *B. cenocepacia* several components important for biofilm formation have been identified[60,138,139]. Collagen-like proteins are also present during this process, as have been demonstrated in *S. pyogenes* and *B. amyloliquefaciens* (Gram-positive)[140,141] and in *Legionella pneumophila* (Gram-negative)[142,143]. In this study, we assess biofilm formation in static conditions for 24 and 48 hours on polystyrene surfaces. The results presented in Figure 14. show that there is no significant difference between mutant and wild-type regarding biofilm formation, either at 24 or 48 hours.

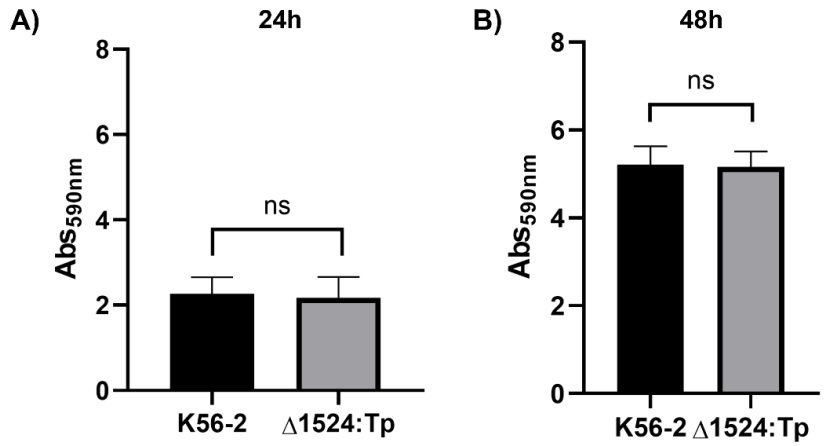


Figure 14. Biofilm formation in polystyrene microtiter plates by wild-type B. cenocepacia K56-2 and the Δ1524::Tp mutant at **A)** 24 hours and **B)** 48 hours. Results are shown as the mean values with error bars representing the standard deviation of three independent replicates with at least 6 samples in each experiment. No difference is observed in biofilm formation ability in the tested conditions (ns, $P > 0.05$).

In *L. pneumophila* was shown that CLPs were important for biofilm formation[143], with the number of repeats of the GXY triplet influencing the robustness of the biofilm[142]. Besides the BCAL1524 protein, which possesses 18 GXY repetitions, *B. cenocepacia* possesses still proteins BCAM0695 and BCAM1598, two other collagen-like proteins with a higher number of GXY repetitions, with *bcam0695* being one of the genes overexpressed in Pimenta et al.[29] transcriptome analysis in adhesion conditions, which may indicate its role in adhesion processes. The presence of these other proteins with similar characteristics and a higher number of GXY repetitions may compensate for the deletion of the *bcal1524* gene.

3.6. Adhesion to epithelial lung cells and to ECM components is not affected

B. cenocepacia adhesion has been studied in the last years, being identified several components essential for this process[62,119,144]. Moreover, Pimenta and colleagues identified the *bcal1524* gene as overexpressed during an adhesion assay[29]. Collagen-like proteins are already described as important for the adhesion process in other organisms, like *B. amyloliquefaciens* to plant roots[141] or in *S. pyogenes* to respiratory epithelial cells[76]. Since the *bcal1524* gene was overexpressed in an adhesion assay and there are other CLPs already described as being part of the adhesion process in other organisms, we performed an adhesion assay to human bronchial epithelial cells (16HBE14o- cell line). As is represented in Figure 15, there is a significant difference in the percentage of adhered bacteria comparing the mutant and the wild-type, with an increasing percentage of mutant bacteria adhering to the host cells. As referred, the adhesion process is multifactorial, involving many proteins of different natures. It is plausible that the existence of several other proteins could mask the deletion of our gene of interest, with other proteins enhancing its activity, which would lead to the augmented adhesion capacity of the mutant and the wild-type. Additionally, the deletion of the BCAL1524 protein may alter the spatial composition of the membrane, exposing other adhesins in a higher level than in the wild-type strain. However, to fully understand the importance of this protein in the adhesion to host cells, more assays would be necessary.

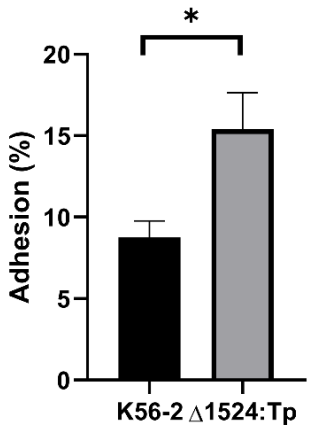


Figure 15. Adherence to 16HBE14o- epithelial cell line by the Δ1524::Tp mutant expressed as the percentage of adhered bacteria relative to the initial dose of bacteria applied to the monolayer of epithelial cells. Results are presented as the mean values with error bars from three independent experiments. A significative difference is observed between the wild-type and the mutant (*, $P < 0.05$).

Furthermore, we proceeded with adhesion assays to immobilized components of the extracellular matrix. The ECM is present in all tissues, with collagen being one of the most abundant proteins in mammalian ECM[145]. The components of the ECM are important targets for pathogens aiming to infect the host, with several proteins described as fibronectin or collagen-binding, from various organisms[146,147]. Collagen-like proteins, like Scl1, were also described as having domains with a binding capacity to some components of the ECM[148]. For that reason, in this study, we analyzed the binding ability of the wild-type and mutant strain to three different components of the ECM: collagen type I, collagen type IV and fibronectin. As observed for the adhesion to the epithelial cells, there were no registered significant differences in the binding capacity of the mutant to none of the tested ECM components (Figure 16).

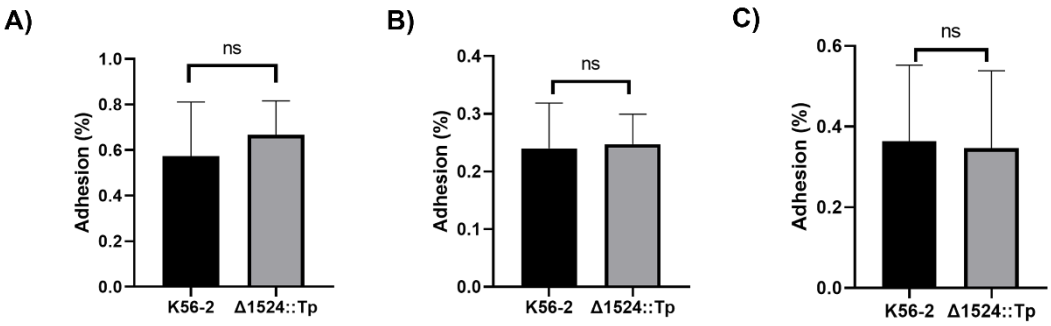


Figure 16 Effect of the Δ1524::Tp mutation on adherence to ECM components: **A)** Collagen type IV, **B)** Fibronectin and **C)** Collagen type I. The results are represented as the mean values with error bars representing SD from three independent experiments. No significant differences were registered for any of the extracellular matrix components (ns, $P > 0.05$).

3.7. Assessing the virulence of the mutant with *Galleria mellonella* larvae

To analyze if the deletion of the *bcal1524* would interfere with the virulence potential of *B. cenocepacia*, a survival assay was performed with larvae from the insect model *Galleria mellonella*. After infection with the mutant or the wild-type strain, the survival of the larvae was assessed. The results expressed in Figure 17 demonstrate that the deletion of the *bcal1524* gene did not affect the killing ability of *B. cenocepacia*.

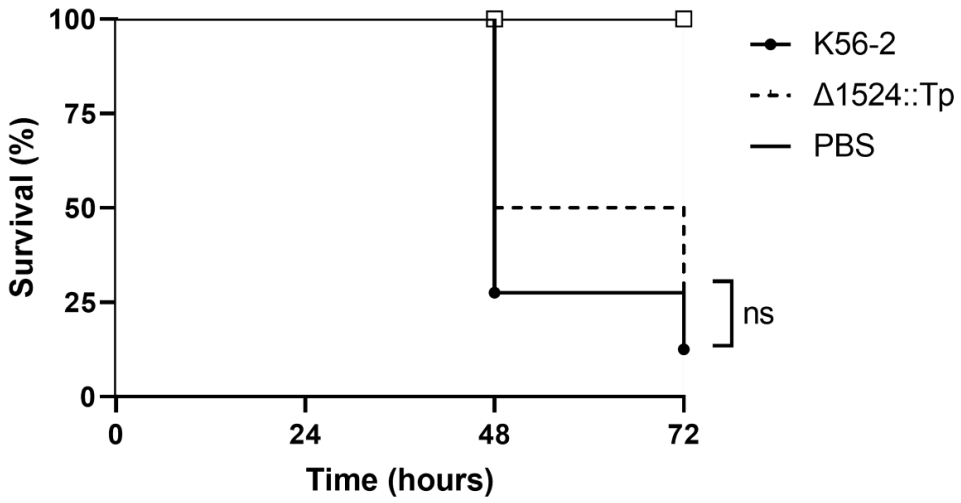


Figure 17. The Δ1524::Tp mutant killing ability was not affected, when compared to the wild-type. Kaplan-Meier graphs of G. mellonella survival after injection with B. cenocepacia K56-2 and B. cenocepacia Δ1524::Tp. Uninfected larvae were injected with PBS were used as control. The results shown represent the mean values of three independent experiments with 10 larvae used per experiment (ns, $P > 0.05$).

3.8. Expression and purification of BCAM0695 and BCAL1524 truncates

In addition to the phenotypic characterization of the mutant, the expression and purification of several truncates from the proteins BCAM0695 and BCAL1524 were attempted. Initially, a fragment with 737 amino acids from the BCAM0695 protein (the entire protein except for the lipid box) (Figure 18) was expressed using the pET28-a plasmid in *E. coli* BL21-DE3, but no significant expression was observable. However, a weak band is observable in a Western blot, indicating the existence of the protein, but not in the quantity required for further purification (data not shown). Thereafter, the construction of four truncates was attempted using the Champion™ pET SUMO Protein Expression System (Invitrogen, Thermo Fisher, USA): two for the BCAM0695, and two for the BCAL1524. This expression system uses the SUMO protein to increase solubility of the expressed protein and allows the further purification adding a 6xHis tag. For both proteins, the objective was to obtain a truncate of the entire protein except for the lipid box (BCAM0695_21_758 and BCAL1524_33_558) and a truncate without the GXY collagen typical repetition domain (BCAM0695_320_758 and BCAL1524_116_558) (Figure 18).

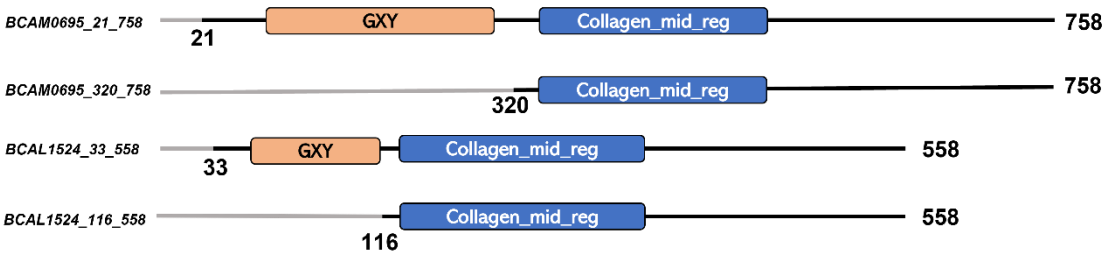


Figure 18. Representation of the truncates designed for expression and purification assays. One with the GXY repeats domain and one without for BCAM0695 (up) and BCAL1524 proteins (down). The missing part of each protein is represented by a grey line.

The constructs were confirmed by sequencing and both constructions of the BCAM0695 protein possessed mutations, whereas both constructs from BCAL1524 were correct. For that reason, the experiment proceeded only with the truncates from BCAL1524. The estimated protein weight for each truncate was of 60 kDa for the BCAL1524_33_558 and of 53 kDa for the BCAL1524_116_558. The next step was to test the optimal expression conditions for each truncate. For that, the constructions were inserted into *E. coli* BL21-DE3 and the overexpression was tested with the following conditions: induction with IPTG (0.2, 0.6 and 1 mM) for 4 hours at 37 ºC, 250 rpm. For each time point and concentration of IPTG a sample of cells was harvested and visualized in a 10% SDS-Page gel (Figure 19). As represented in the figure, we observed that for the BCAL1524_33_558 truncate the optimal time of incubation was 4 hours, with an optimal concentration of IPTG of 0.6 mM (Figure 19 A)). For the truncate BCAL1524_116_558, the optimal incubation time was also 4 hours, but with a needed concentration of 1 mM of IPTG (Figure 19 B)).

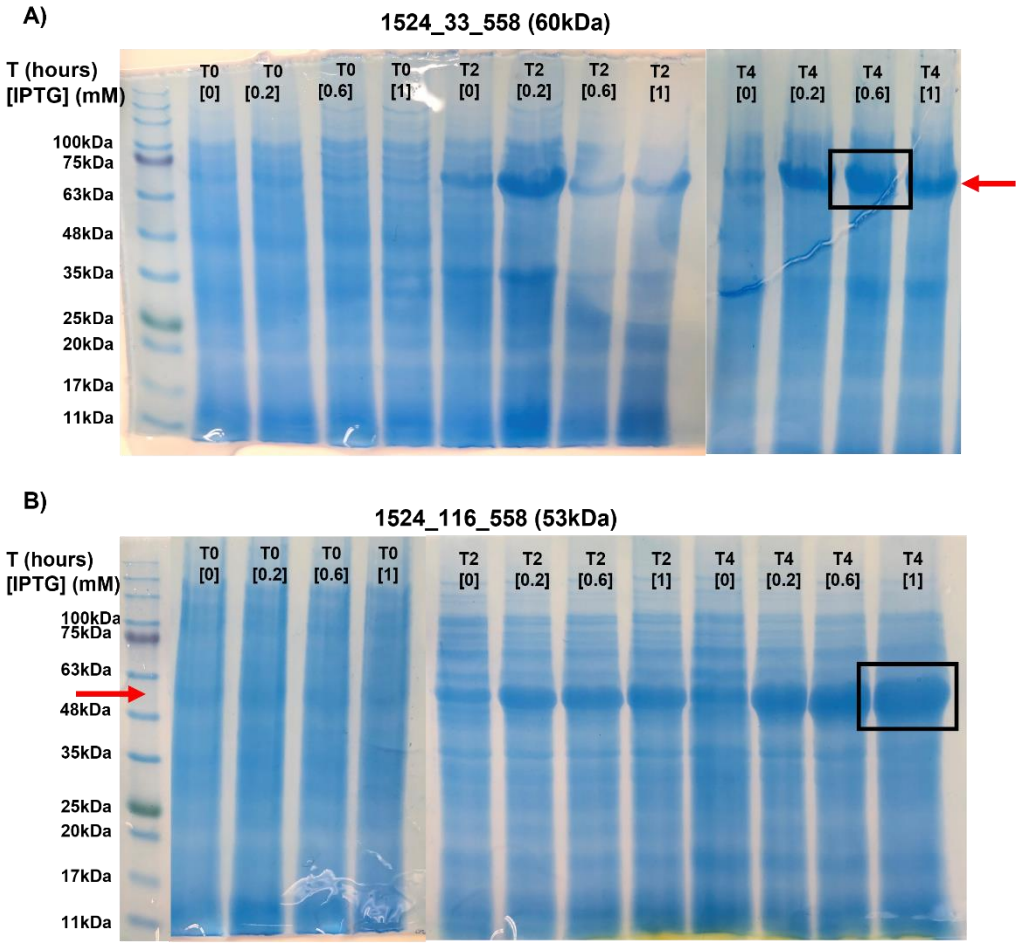


Figure 19. SDS-Page gels (10%) with extracts from each condition of expression tested for both truncates. **A)** for BCAL1524_33_558, **B)** for BCAL1524_116_558. Red arrows indicate the size of the truncates. Boxes highlight the optimal conditions of expression for each truncate.

After identifying the optimal conditions, an expression and further purification were performed. Afterwards, the aim was to test the optimal conditions for the cleavage of the SUMO protein from our truncates of interest, to obtain the purified native proteins. SUMO protease was incubated with our truncates of interest for 1 hour at 30 ºC. Furthermore, a purification was performed to separate the cleaved SUMO protein from the native truncates. In Figure 20. it is observable the efficient cleavage of the SUMO protein, which allowed the collection and storage of the native truncates, even though retaining some contamination. Thus, we identified that the temperature of 37 ºC was suitable for the expression of the truncates, with the time of incubation and concentration of IPTG being specific for each one (Figure 19, black boxes). We were also able to confirm that the tested conditions allowed the cleavage of the SUMO protein at the end of the protocol, allowing the recovery of the native truncates, albeit with some contamination with other proteins.

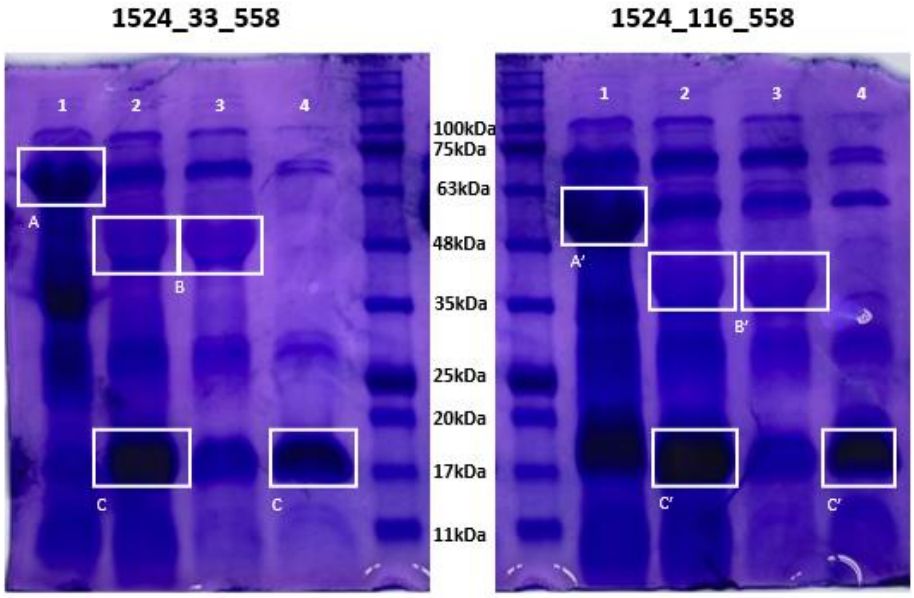


Figure 20. SDS-Page gel (10%) with extracts of the cleavage of the SUMO from the truncates BCAL1524_33_558 and BCAL1524_116_558. 1) Truncate-SUMO; 2) Native Truncate + SUMO; 3) Purified Native Truncate; 4) SUMO protein. A and A') Truncate-SUMO before cleavage of the two truncates, respectively; B and B') Native Truncates, respectively; C and C') SUMO protein after cleavage, respectively.

4. Conclusions and Future Perspectives

Pimenta et al. performed a transriptomic analysis during the first contact between *B. cenocepacia* K56-2 and human bronchial cells and identified that several collagen-like genes were present as overexpressed[29]. The aim of this work was then to study the existence and the characteristics of this collagen-like proteins in the Bcc. From the nineteen members of the Bcc, 75 proteins with the Bacterial Collagen Middle Region (Col_mid_reg) were identified. From those, 93% possess tandem repeats, a characteristic associated with pathogenicity[130,131]. Analysing the distribution per species, there seems to be a tendency for an increased percentage of tandem repeats and a higher level of pathogenicity. However, more studies are necessary to increase the understanding of these regions. Furthermore, we analyzed the cellular localization and nature of the proteins, concluding that the majority are in the cytoplasmic or extracellular membrane, and being described as lipoproteins. Focusing on the proteins from *B. cenocepacia*, a low complexity and disorder analysis showed a great prevalence of disordered regions in the 5 proteins, with these regions possessing high nucleic acid and proteins binding affinity, important for adhesion and infection processes[90,91]. These proteins, possessing low complexity regions and the Col_mid_reg, may suggest the existence of a different type of CLPs. Besides the CLPs possessing GXY domains that originate triple helixes and structures proteins, this unstructured proteins with a common central domain may possess interesting characteristics yet to be studied. This type of discoveries, adding to the existing evidence of an overexpression of the genes *bcal1523*, *bcam0695* and *bcal1524* in adhesion assays[29] led us to creating mutants in order to understand the eventual role and importance of these protein during the adhesion process, assessing several steps and events that are key to the pathogenicity associated to *B. cenocepacia*. Unfortunately, we were only able to create a mutant for the *bcal1524* gene. A phenotypic analysis was performed with this mutant, and the only impaired functions in the mutant when compared to the wild type *B. cenocepacia* K56-2 strain was motility (swimming and swarming were impaired) and the adhesion capacity to host cells, that was augmented in the mutant. This may indicate the importance of this gene for flagella correct formation or function. The increase in adhesion capacity may be explained by the multitude of adhesins presented by the bacteria, specially the other 4 CLPs paralogs identified that may mask the absence of the BCAL1524. The possible alteration in the spatial composition of the membrane of the mutant is another hypothesis to explain this increase in the adhesion to human bronchial cells. The other assays, such as biofilm formation, adhesion to components of the ECM, or even the killing ability were not altered. This can be explained by the fact that these processes are usually dependent of several components, with the absence of just one being not enough to completely impair these processes. As mentioned before, the existence of four other CLPs with similar characteristics may result in a redundant activity that can compensate the deletion of one specific protein. However, more studies are necessary to fully understand the importance of this protein. For that reason, protein truncates of the BCAL1524 protein were successfully expressed and purified. Further studies to assess the function of each domain of the protein may be important to understand the structure, importance, and eventual potential as a target to tackle this opportunistic pathogen. Additionally, more attempts to construct the *bcam0695* mutant and the BCAM0695 mutants is necessary, since this will allow the study of the importance of GXY repeats (present in high number in this protein) and the collagen-like domains in the virulence of *B. cenocepacia*,

namely in adherence and invasion to the host cells or biofilm formation, as is already described in other organisms[68,76,77].

5. References

1. Dunne, W. M. Bacterial adhesion: Seen any good biofilms lately? *Clin. Microbiol. Rev.* **15**, 155–166 (2002).

2. Donlan, R. M. Biofilm formation: A clinically relevant microbiological process. *Clin. Infect. Dis.* **33**, 1387–1392 (2001).

3. Notermans, S., Dormans, J. A. M. A. & Mead, G. C. Contribution of surface attachment to the establishment of micro-organisms in food processing plants: A review. *Biofouling* **5**, 21–36 (1991).

4. Cooksey, K. E. & Wigglesworth-Cooksey, B. Adhesion of bacteria and diatoms to surfaces in the sea: A review. *Aquat. Microb. Ecol.* **9**, 87–96 (1995).

5. Cossart, P. & Pizarro-cerda, J. Bacterial Adhesion and Entry into Host Cells. *Cell* **124**, 715–727 (2006).

6. Tuson, Hannah H.; Weibel, D. B. Bacteria-surface interactions. *Soft Matter* **9**, 4368–4380 (2013).

7. Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. & Lappin-Scott, H. M. Microbial biofilms. *Annu. Rev. Microbiol.* **49**, 711–745 (1995).

8. Berne, C., Ellison, C. K., Ducret, A. & Brun, Y. V. Bacterial adhesion at the single-cell level. *Nat. Rev. Microbiol.* **16**, 616–627 (2018).

9. Carpentier, B. & Cerf, O. Biofilms and their consequences, with particular reference to hygiene in the food industry. *J. Appl. Bacteriol.* **75**, 499–511 (1993).

10. Garrett, T. R., Bhakoo, M. & Zhang, Z. Bacterial adhesion and biofilms on surfaces. *Prog. Nat. Sci.* **18**, 1049–1056 (2008).

11. Kumar, C. G. & Anand, S. K. Significance of microbial biofilms in food industry: A review. *Int. J. Food Microbiol.* **42**, 9–27 (1998).

12. Dickson, J. S. & Koohmaraie, M. Cell surface charge characteristics and their relationship to bacterial attachment to meat surfaces. *Appl. Environ. Microbiol.* **55**, 832–836 (1989).

13. Palmer, J., Flint, S. & Brooks, J. Bacterial cell attachment, the beginning of a biofilm. *J. Ind. Microbiol. Biotechnol.* **34**, 577–588 (2007).

14. Marshall, K. C., Stout, R. & Mitchell, R. Mechanism of the Initial Events in the Sorption of Marine Bacteria to Surfaces. *J. Gen. Microbiol.* **68**, 337–348 (1971).

15. Boks, N. P., Norde, W., van der Mei, H. C. & Busscher, H. J. Forces involved in bacterial adhesion to hydrophilic and hydrophobic surfaces. *Microbiology* **154**, 3122–3133 (2008).

16. Ren, Y. *et al.* Emergent heterogeneous microenvironments in biofilms : substratum surface heterogeneity and bacterial adhesion force-sensing. *FEMS Microbiol. Rev.* **42**, 259–272 (2018).

17. Camesano, T. A. & Logan, B. E. Probing bacterial electrosteric interactions using atomic force microscopy. *Environ. Sci. Technol.* **34**, 3354–3362 (2000).

18. Renner, L. D. & Weibel, D. B. Physicochemical regulation of biofi lm formation. *MRS Bulletin* **36**, 347–355 (2011).

19. An, Y. H. & Friedman, R. J. Concise review of mechanisms of bacterial adhesion to biomaterial surfaces. *J. Biomed. Mater. Res.* **43**, 338–348 (1998).

20. Berne, C., Ducret, A., Hardy, G. G. & Brun, Y. V. Adhesins Involved in Attachment to Abiotic Surfaces by Gram-Negative Bacteria. *Microbiol. Spectr.* **3**, 1–27 (2015).

21. Alam, F., Kumar, S. & Varadarajan, K. M. Quantification of Adhesion Force of Bacteria on the Surface of Biomaterials: Techniques and Assays. *ACS Biomater. Sci. Eng.* **5**, 2093–2110 (2019).

22. Reyes, C. D. & Garcı, J. A centrifugation cell adhesion assay for high-throughput screening of biomaterial surfaces. *J. Biomed. Mater. Res. - Part A* **67**, 328-333 (2003).

23. Thewes, N. *et al.* Hydrophobic interaction governs unspecific adhesion of staphylococci : a single cell force spectroscopy study. *Beilstein J. Nanotechnol.* **5**, 1501–1512 (2014).

24. Beaussart, A. *et al.* Single-cell force spectroscopy of probiotic bacteria. *Biophys. J.* **104**, 1886–1892 (2013).

25. El-Kirat-Chatel, S., Mil-Homens, D., Beaussart, A., Fialho, A. M. & Dufrêne, Y. F. Single-molecule atomic force microscopy unravels the binding mechanism of a *Burkholderia cenocepacia* trimeric autotransporter adhesin. *Mol. Microbiol.* **89**, 649–659 (2013).

26. Williamson, Y. M. *et al.* Surfaceome Analysis Protocol for the Identification of Novel *Bordetella pertussis* Antigens. in **1722**, 3–20 (2018).

27. Khater, F. *et al.* In silico analysis of usher encoding genes in *Klebsiella pneumoniae* and characterization of their role in adhesion and colonization. *PLoS One* **10**, 1–24 (2015).

28. Bierne, H. & Cossart, P. *Listeria monocytogenes* Surface Proteins: from Genome Predictions to Function . *Microbiol. Mol. Biol. Rev.* **71**, 377–397 (2007).

29. Pimenta, A. I., Bernardes, N., Alves, M. M. & Homens, D. M. *Burkholderia cenocepacia* transcriptome during the early contacts with giant plasma membrane vesicles derived from live bronchial epithelial cells. *Sci. Rep.* **11**, 1–16 (2021).

30. Beydokhti, S. S., Stork, C., Dobrindt, U. & Hensel, A. *Orthosipon stamineus* extract exerts inhibition of bacterial adhesion and chaperon-usher system of uropathogenic *Escherichia coli*—a transcriptomic study. *Appl. Microbiol. Biotechnol.* **103**, 8571–8584 (2019).

31. Sackstein, R. & Fuhlbrigge, R. The Blot Rolling Assay. **341**, 217–226

32. Haiko, J. & Westerlund-Wikström, B. The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel).* **2**, 1242–1267 (2013).

33. Cairns, L. S., Marlow, V. L., Bissett, E., Ostrowski, A. & Stanley-Wall, N. R. A mechanical signal transmitted by the flagellum controls signalling in Bacillus subtilis. *Mol. Microbiol.* **90**, 6–21 (2013).

34. Girón, J. A., Torres, A. G., Freer, E. & Kaper, J. B. The flagella of enteropathogenic Escherichia coli mediate adherence to epithelial cells. *Mol. Microbiol.* **44**, 361–379 (2002).

35. Kreis, C. T., Grangier, A. & Bäumchen, O. In vivo adhesion force measurements of Chlamydomonas on model substrates. *Soft Matter* **15**, 3027–3035 (2019).

36. Smyth, C. J., Marron, M. B., Twohig, J. M. G. J. & Smith, S. G. J. Fimbrial adhesins : similarities and variations in structure and biogenesis. **16**, 127–139 (1996).

37. Schroll, C., Barken, K. B., Krogfelt, K. A. & Struve, C. Role of type 1 and type 3 fimbriae in Klebsiella pneumoniae biofilm formation. *BMC Microbiol.* **10**, 179 (2010).

38. Foroogh, N., Rezvan, M., Ahmad, K. & Mahmood, S. Structural and functional characterization of the FimH adhesin of uropathogenic *Escherichia coli* and its novel applications. *Microb. Pathog.* **161**, 105288 (2021).

39. Rehman, T. *et al.* Microbial Pathogenesis Adhesive mechanism of different Salmonella fimbrial adhesins. *Microb. Pathog.* **137**, 103748 (2019).

40. Hori, K. & Matsumoto, S. Bacterial adhesion: From mechanism to control. *Biochem. Eng. J.* **48**, 424–434 (2010).

41. Pratt, L. A. & Kolter, R. Genetic analysis of Escherichia coli biofilm formation: roles of flagella, motility, chemotaxis and type I pili. **30**, 285–293 (1998).

42. Mack, D. *et al.* Essential functional role of the polysaccharide intercellular adhesin of Staphylococcus epidermidis in hemagglutination. *Infect. Immun.* **67**, 1004–1008 (1999).

43. Dramsi, S. *et al.* Assembly and role of pili in group B streptococci. *Mol. Microbiol.* **60**, 1401–1413 (2006).

44. Chagnot, C., Zorgani, M. A., Astruc, T. & Desvaux, M. Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. *Front. Microbiol.* **4**, 1–26 (2013).

45. Foster, T. J. The remarkably multifunctional fibronectin binding proteins of *Staphylococcus aureus. Eur. J. Clin. Microbiol. Infect. Dis.* **35**, 1923–1931 (2016).

46. Upreti, R. K. & Kumar, M. Review Bacterial glycoproteins : Functions, biosynthesis and applications. *Proteomics* **3**, 363–379 (2003).

47. Tytgat, H. L. P. & de Vos, W. M. Sugar Coating the Envelope: Glycoconjugates for Microbe–Host Crosstalk. *Trends Microbiol.* **24**, 853–861 (2016).

48. Balonova, L. *et al.* Characterization of Protein Glycosylation in Francisella tularensis subsp . holarctica. *Mol. Cell. Proteomics* **11**, M111.015016-1-M111.015016-12 (2012).

49. Charbonneau, M. È., Girard, V., Nikolakakis, A., Campos, M. & Dumas, F. O-Linked Glycosylation Ensures the Normal Conformation of the Autotransporter Adhesin Involved in Diffuse Adherence. *J. Bacteriol.* **189**, 8880–8889 (2007).

50. Hanuszkiewicz, A. *et al.* Identification of the Flagellin Glycosylation System in Burkholderia cenocepacia and the Contribution of Glycosylated Flagellin to Evasion of Human Innate Immune Responses. *J. Biol. Chem.* **289**, 19231–19244 (2014).

51. Kostakioti, M., Newman, C. L., Thanassi, D. G. & Stathopoulos, C. Mechanisms of protein export across the bacterial outer membrane. *J. Bacteriol.* **187**, 4306–4314 (2005).

52. Meuskens, I., Saragliadis, A., Leo, J. C. & Linke, D. Type V secretion systems: An overview of passenger domain functions. *Front. Microbiol.* **10**, 1–19 (2019).

53. Benz, I. & Schmidt, M. A. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic Escherichia coli. *Infect. Immun.* **57**, 1506–1511 (1989).

54. Laarmann, S. & Schmidt, M. A. The Escherichia coli AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor. *Microbiology* **149**, 1871–1882 (2003).

55. Braun, D. *et al.* Receptor Analogs and Monoclonal Antibodies that Inhibit Adherence of Bordetella pertussis to Human Ciliated Respiratory Epithelial Cells. *J. Exp. Med.* **168**, 267–277 (1988).

56. Cotter, S. E., Surana, N. K. & St. Geme, J. W. Trimeric autotransporters: A distinct subfamily of autotransporter proteins. *Trends Microbiol.* **13**, 199–205 (2005).

57. Girard, V. & Mourez, M. Adhesion mediated by autotransporters of Gram-negative bacteria: Structural and functional features. *Res. Microbiol.* **157**, 407–416 (2006).

58. Tahir, Y. El & Skurnik, M. YadA , the multifaceted Yersinia adhesin. *Int. J. Med. Microbiol.* **218**, 209–218 (2001).

59. Mil-Homens, D., Rocha, E. P. C. & Fialho, A. M. Genome-wide analysis of DNA repeats in Burkholderia cenocepacia J2315 identifies a novel adhesin-like gene unique to epidemic-associated strains of the ET-12 lineage. *Microbiology* **156**, 1084–1096 (2010).

60. Mil-Homens, D., Leç, M. I., Fernandes, F., Pinto, S. N. & Fialho, A. M. Characterization of BCAM0224, a multifunctional trimeric autotransporter from the human pathogen Burkholderia cenocepacia. *J. Bacteriol.* **196**, 1968–1979 (2014).

61. Pimenta, A. I., Mil-Homens, D. & Fialho, A. M. Burkholderia cenocepacia–host cell contact controls the transcription activity of the trimeric autotransporter adhesin BCAM2418 gene. *Microbiologyopen* **9**, 1–12 (2020).

62. Pimenta, A. I., Mil-homens, D. & Kilcoyne, M. Burkholderia cenocepacia BCAM2418-induced antibody inhibits bacterial adhesion, confers protection to infection and enables identification of host glycans as adhesin targets. *Cell. Microbiol.* **23**, 1–16 (2021).

63. Duarte, A. S., Correia, A. & Esteves, A. C. Bacterial collagenases - A review. *Crit. Rev. Microbiol.*

42, 106–126 (2016).

64. Brodsky, B. & Ramshaw, J. A. M. The collagen triple-helix structure. *Matrix Biol.* **15**, 545–554 (1997).

65. Brodsky, B. & Persikov, A. V. Molecular structure of the collagen triple helix. *Adv. Protein Chem.* **70**, 301–339 (2005).

66. Rasmussen, M., Jacobsson, M. & Bjo, L. Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins. *J. Biol. Chem.* **278**, 32313–32316 (2003).

67. Engel, J., Peter, H. & Bächinger, L. Collagen-like sequences in phages and bacteria. *Proc. Indian Acad. Sci. Chem. Sci.* **111**, 81–86 (1999).

68. Yu, Z., An, B., Ramshaw, J. A. M. & Brodsky, B. Bacterial collagen-like proteins that form triple-helical structures. *J Struct Biol* **186**, 451–461 (2014).

69. Brodsky, B. & Ramshaw, J. A. M. Bioengineered collagens. *Subcell. Biochem.* **82**, 601–629 (2017).

70. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. *Annu. Rev. Microbiol.* **55**, 709–742 (2001).

71. Lukomski, S., Bachert, B. A., Squeglia, F. & Berisio, R. Collagen-like proteins of pathogenic streptococci. *Mol. Microbiol.* **103**, 919–930 (2017).

72. McElroy, K., Mouton, L., Du Pasquier, L., Qi, W. & Ebert, D. Characterisation of a large family of polymorphic collagen-like proteins in the endospore-forming bacterium Pasteuria ramosa. *Res. Microbiol.* **162**, 701–714 (2011).

73. Doxey, A. C. & Mcconkey, B. J. Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* **4**, 453–466 (2013).

74. Bella, J. A first census of collagen interruptions: Collagen's own stutters and stammers. *J. Struct. Biol.* **186**, 438–450 (2014).

75. Paterson, G. K., Nieminen, L., Jefferies, J. M. C. & Mitchell, T. J. PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol. Lett.* **285**, 170–176 (2008).

76. Chen, S. M. *et al.* Streptococcal collagen-like surface protein 1 promotes adhesion to the respiratory epithelial cell. *BMC Microbiol.* **10**, 320 (2010).

77. Xu, Y., Keene, D. R., Bujnicki, J. M. & Ho, M. Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices. **277**, 27312–27318 (2002).

78. Lukomski, S. *et al.* Identification and characterization of the scl gene encoding a group A Streptococcus extracellular protein virulence factor with similarity to human collagen. *Infect. Immun.* **68**, 6542–6553 (2000).

79. Rasmussen, M., Ede, A. & Bjo, L. SclA , a Novel Collagen-Like Surface Protein of Streptococcus pyogenes. *Infect. Immun.* **68**, 6370–6377 (2000).

80. Bachert, B. A., Choi, S. J., Snyder, A. K., Rio, R. V. M. & Durney, B. C. A Unique Set of the Burkholderia Collagen- Like Proteins Provides Insight into Pathogenesis , Genome Evolution and Niche Adaptation , and Infection Detection. *PLoS One* **10**, 1–36 (2015).

81. Grund, M. E. *et al.* Burkholderia collagen-like protein 8, Bucl8, is a unique outer membrane component of a putative tetrapartite efflux pump in Burkholderia pseudomallei and Burkholderia mallei. *PLoS One* **15**, 1–26 (2020).

82. Qiu, Y., Zhai, C., Chen, L., Liu, X. & Yeo, J. Current Insights on the Diverse Structures and Functions in Bacterial Collagen-like Proteins. *ACS Biomater. Sci. Eng.* (2021).

83. Peng, Y. Y. *et al.* Towards scalable production of a collagen-like protein from Streptococcus pyogenes for biomedical applications. *Microb. Cell Fact.* **11**, 1–8 (2012).

84. Bronk, J. K. *et al.* A multifunctional streptococcal collagen-mimetic protein coating prevents bacterial adhesion and promotes osteoid formation on titanium. *Acta Biomater.* **10**, 3354–3362 (2014).

85. Cereceres, S. *et al.* Chronic Wound Dressings Based on Collagen-Mimetic Proteins. *Adv. Wound Care* **4**, 444–456 (2015).

86. Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **114**, 6561–6588 (2014).

87. Campen, A. *et al.* TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder. *Protein Pept. Lett.* **15**, 956–963 (2008).

88. Coletta, A. *et al.* Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* **4**, (2010).

89. Chen, J. W., Romero, P., Uversky, V. N. & Dunker, A. K. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J. Proteome Res.* **5**, 879–887 (2006).

90. Xue, B. *et al.* Orderly order in protein intrinsic disorder distribution : disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **30**, 137-149 (2012).

91. Peng, Z. *et al.* Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **72**, 137–151 (2015).

92. Mier, P. & Andrade-Navarro, M. A. The conservation of low complexity regions in bacterial proteins depends on the pathogenicity of the strain and subcellular location of the protein. *Genes (Basel).* **12**, (2021).

93. Krachler, A. M. & Orth, K. Targeting the bacteria-host interface strategies in anti-adhesion therapy. *Virulence* **4**, 284–294 (2013).

94. Wojnicz, D., Klaka, M., Adamski, R. & Jankowski, S. Influence of Subinhibitory Concentrations of Amikacin and Ciprofloxacin on Morphology and Adherence Ability of Uropathogenic Strains. *Folia Microbiol* **52**, 429–436 (2007).

95. Radin, N. S. Preventing the binding of pathogens to the host by controlling sphingolipid metabolism. *Microbes Infect.* **8**, 938–945 (2006).

96. Svensson, M. *et al.* Glycolipid depletion in antimicrobial therapy. *Mol. Microbiol.* **47**, 453–461 (2003).

97. Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta - Gen. Subj.* **1760**, 527–537 (2006).

98. Mydock-McGrane, L. K., Hannah, T. J. & Janetka, J. W. Rational Design Strategies for FimH Antagonists: New Drugs on the Horizon for Urinary Tract Infection and Crohn's Disease. *Expert Opin Drug Discov* **12**, 711–731 (2017).

99. Lee, K. K. *et al.* Use of synthetic peptides in characterization of microbial adhesins. *Methods Enzymol.* **253**, 115–131 (1995).

100. Ghosh, S. *et al.* An adhesion protein of Salmonella enterica serovar Typhi is required for pathogenesis and potential target for vaccine development. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3348–3353 (2011).

101. Ofek, I., Hasty, D. L. & Sharon, N. Anti-adhesion therapy of bacterial diseases: Prospects and problems. *FEMS Immunol. Med. Microbiol.* **38**, 181–191 (2003).

102. Allen, R. C., Popat, R., Diggle, S. P. & Brown, S. P. Targeting virulence: Can we make evolution-proof drugs? *Nat. Rev. Microbiol.* **12**, 300–308 (2014).

103. Ternent, L., Dyson, R. J., Krachler, A. & Jabbari, S. Bacterial fitness shapes the population dynamics of antibiotic-resistant and -susceptible bacteria in a model of combined antibiotic and anti-virulence treatment. *J. Theor. Biol.* **372**, 1–11 (2015).

104. Stones, D. H. & Krachler, A. M. Against the tide : the role of bacterial adhesion in host colonization. *Biochem. Soc. Trans.* **44**, 1571–1580 (2016).

105. Coenye, T., Vandamme, P., Govan, J. R. W. & Lipuma, J. J. Taxonomy and identification of the Burkholderia cepacia complex. *J. Clin. Microbiol.* **39**, 3427–3436 (2001).

106. Mahenthiralingam, E., Urban, T. A. & Goldberg, J. B. The multifarious, multireplicon Burkholderia cepacia complex. *Nat. Rev. Microbiol.* **3**, 144–156 (2005).

107. Lessie, T. G., Hendrickson, W., Manning, B. D. & Devereux, R. Genomic complexity and plasticity of Burkholderia cepacia. *FEMS Microbiol. Lett.* **144**, 117–128 (1996).

108. Goldmann, D. A. & Klinger, J. D. Pseudomonas cepacia: Biology, mechanisms of virulence, epidemiology. *J. Pediatr.* **108**, 806–812 (1986).

109. Isles, A. *et al.* Pseudomonas cepacia infection in cystic fibrosis: An emerging problem. *J. Pediatr.* **104**, 206–210 (1984).

110. Biddick, R., Spilker, T., Martin, A. & LiPuma, J. J. Evidence of transmission of Burkholderia cepacia, Burkholderia multivorans and Burkholderia dolosa among persons with cystic fibrosis. *FEMS Microbiol. Lett.* **228**, 57–62 (2003).

111. Mahenthiralingam, E., Baldwin, A. & Vandamme, P. Burkholderia cepacia complex infection in patients with cystic fibrosis. *J. Med. Microbiol.* **51**, 533–538 (2002).

112. Johnson, W. M., Tyler, S. D. & Rozee, K. R. Linkage analysis of geographic and clinical clusters in Pseudomonas cepacia infections by multilocus enzyme electrophoresis and ribotyping. *J. Clin. Microbiol.* **32**, 924–930 (1994).

113. Drevinek, P. & Mahenthiralingam, E. Burkholderia cenocepacia in cystic fibrosis: Epidemiology and molecular mechanisms of virulence. *Clin. Microbiol. Infect.* **16**, 821–830 (2010).

114. McClean, S. & Callaghan, M. Burkholderia cepacia complex: Epithelial cell-pathogen confrontations and potential for therapeutic intervention. *J. Med. Microbiol.* **58**, 1–12 (2009).

115. Tomich, M., Herfst, C. A., Golden, J. W. & Mohr, C. D. Role of Flagella in Host Cell Invasion by Burkholderia cepacia. *Infect. Immun.* **70**, 1799–1806 (2002).

116. McClean, S. *et al.* Linocin and OmpW are involved in attachment of the cystic fibrosis-associated pathogen Burkholderia cepacia complex to lung epithelial cells and protect mice against infection. *Infect. Immun.* **84**, 1424–1437 (2016).

117. Dennehy, R. *et al.* The Burkholderia cenocepacia peptidoglycan-associated lipoprotein is involved in epithelial cell attachment and elicitation of inflammation. *Cell. Microbiol.* **19**, (2017).

118. Mil-Homens, D. & Fialho, A. M. Trimeric autotransporter adhesins in members of the Burkholderia cepacia complex: A multifunctional family of proteins implicated in virulence. *Front. Cell. Infect. Microbiol.* **1**, 1–10 (2011).

119. Mil-Homens, D. & Fialho, A. M. A BCAM0223 mutant of Burkholderia cenocepacia is deficient in hemagglutination, serum resistance, adhesion to epithelial cells and virulence. *PLoS One* **7**, (2012).

120. Goldberg, J. B., Ganesan, S., Comstock, A. T., Zhao, Y. & Sajjan, U. S. Cable pili and the associated 22 kDa adhesin contribute to Burkholderia cenocepacia persistence in vivo. *PLoS One* **6**, 1–11 (2011).

121. Winsor, G. L. *et al.* The Burkholderia Genome Database : facilitating flexible queries and comparative analyses. *Bioinformatics* **24**, 2803–2804 (2008).

122. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. **26**, 1608–1615 (2010).

123. Lemoine, F. *et al.* NGPhylogeny.fr: New generation phylogenetic services for non-specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019).

124. Newman, A. M. & Cooper, J. B. XSTREAM: A practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics* **8**, 1–19 (2007).

125. Hu, G. *et al.* flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nat. Commun.* **12**, 1–8 (2021).

126. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).

127. Zheng, W. *et al.* Folding non-homologous proteins by coupling deep- learning contact maps with I-TASSER assembly simulations ll Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Reports Methods* **1**, 100014 (2021).

128. Cozens, A. L. *et al.* CFTR expression and chloride secretion in polarized immortal human bronchial epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **10**, 38–47 (1994).

129. Mil-Homens, D., Ferreira-Dias, S. & Fialho, A. M. Fish oils against Burkholderia and Pseudomonas aeruginosa: In vitro efficacy and their therapeutic and prophylactic effects on infected Galleria mellonella larvae. *J. Appl. Microbiol.* **120**, 1509–1519 (2016).

130. Zhou, K., Aertsen, A. & Michiels, C. W. The role of variable DNA tandem repeeats in bacterial adaptation. *FEMS Microbiol. Rev.* **10**, 119–141 (2014).

131. Saravanan, K. M. Sequence and structural analysis of fibronectin - binding protein reveals importance of multiple intrinsic disordered tandem repeats. *J. Mol. Recognit.* **32**, 1–9 (2019).

132. Holden, M. T. G. *et al.* The genome of Burkholderia cenocepacia J2315, an epidemic pathogen of cystic fibrosis patients. *J. Bacteriol.* **91**, 261–277 (2009).

133. Leitão, J. H. *et al.* Pathogenicity, virulence factors, and strategies to fight against Burkholderia cepacia complex pathogens and related species. *Appl. Microbiol. Biotechnol.* **87**, 31–40 (2010).

134. Nykyri, J. *et al.* Role and Regulation of the Flp/Tad Pilus in the Virulence of Pectobacterium atrosepticum SCRI1043 and Pectobacterium wasabiae SCC3193. *PLoS One* **8**, (2013).

135. Alteri, C. J. *et al.* The Flp type IV pilus operon of Mycobacterium tuberculosis is expressed upon interaction with macrophages and alveolar epithelial cells. *Front. Cell. Infect. Microbiol.* **12**, 1–15 (2022).

136. Chaban, B., Hughes, H. V. & Beeby, M. The flagellum in bacterial pathogens: For motility and a whole lot more. *Semin. Cell Dev. Biol.* **46**, 91–103 (2015).

137. Zhao, X., Wang, R., Shang, Q., Hao, H. & Li, Y. The new flagella-associated collagen-like proteins ClpB and ClpC of Bacillus amyloliquefaciens FZB42 are involved in bacterial motility. *Microbiol. Res.* **184**, 25–31 (2016).

138. Bellich, B. *et al.* The biofilm of Burkholderia cenocepacia H111 contains an exopolysaccharide composed of L-rhamnose and L-mannose: Structural characterization and molecular modelling. *Carbohydr. Res.* **499**, 108231 (2021).

139. Wang, K. *et al.* A LysR Family Transcriptional Regulator Modulates Burkholderia cenocepacia Biofilm Formation and Protease Production. *Appl. Environ. Microbiol.* **87**, 1–13 (2021).

140. Oliver-Kozup, H. *et al.* The group A streptococcal collagen-like protein 1, Scl1, mediates biofilm formation by targeting the EDA-containing variant of cellular fibronectin expressed in wounded tissue. *Mol Microbiol* **87**, 672–689 (2013).

141. Zhao, X. *et al.* ClpD) Are Required for Biofilm Formation and Adhesion to Plant Roots by Bacillus amyloliquefaciens FZB42. *PLoS One* **10**, 1–16 (2015).

142. Abdel-Nour, M. *et al.* Polymorphisms of a collagen-like adhesin contributes to legionella pneumophilaadhesion, biofilm formation capacity and clinical prevalence. *Front. Microbiol.* **10**, 1–15 (2019).

143. Mallegol, J. *et al.* Essential Roles and Regulation of the Legionella pneumophila Collagen-Like

Adhesin during Biofilm Formation. *PLoS One* **7**, 1–15 (2012).

144. Seixas, M. M. *et al.* A Polyclonal Antibody Raised against the Burkholderia cenocepacia OmpA-like Protein BCAL2645 Impairs the Bacterium Adhesion and Invasion of Human Epithelial Cells In Vitro. *Biomedicines* **9**, 1788 (2021).

145. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200 (2010).

146. Westerlund, B. & Korhonen, T. K. Bacterial proteins binding to the mammalian extracellular matrix. *Mol. Microbiol.* **9**, 687–694 (1993).

147. Esgleas, M., Lacouture, S. & Gottschalk, M. Streptococcus suis serotype 2 binding to extracellular matrix proteins. *FEMS Microbiol. Lett.* **244**, 33–40 (2005).

148. Caswell, C. C., Oliver-Kozup, H., Han, R., Lukomska, E. & Lukomski, S. Scl1, the multifunctional adhesin of group A Streptococcus selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol. Lett.* **303**, 61–68 (2010).