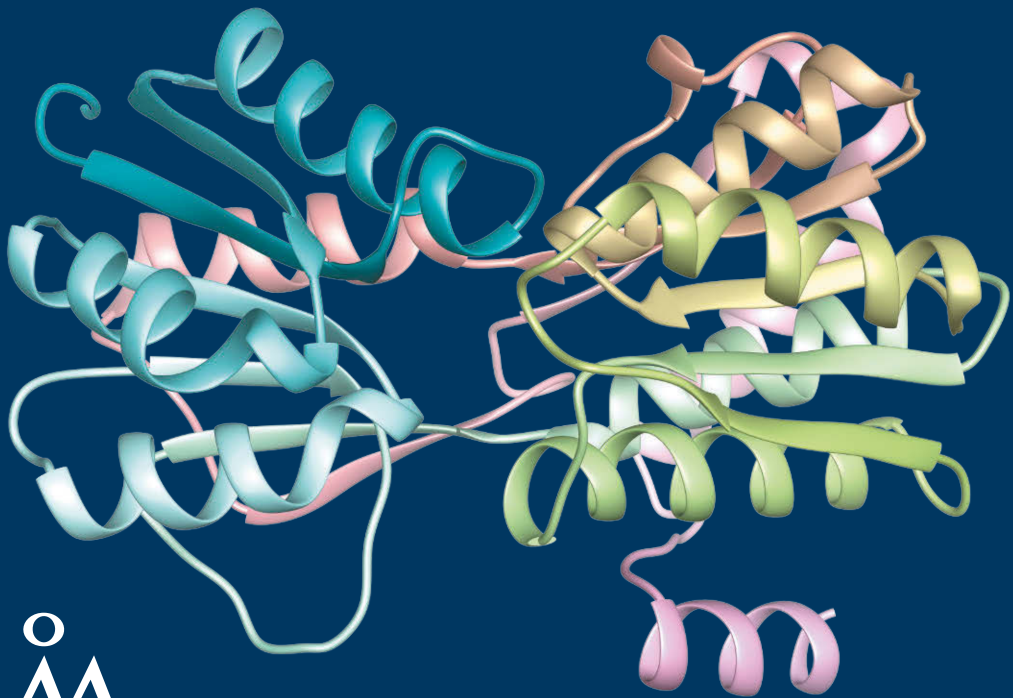
Mia Åstrand

Virulence and survival mechanisms in *Borrelia* and *Klebsiella* infections – a structural bioinformatics perspective



Åbo Akademi
University

Mia Åstrand

Was born on June 2, 1985, in Pietarsaari, Finland. She graduated from University of Helsinki in 2010 with a Master of Science in Biology and obtained a bachelor's degree in Pharmacy in 2016. This PhD thesis project in Biochemistry has taken place during 2016-2021 under the supervision of Adjunct Professor Tiina A. Salminen at the Faculty of Science and Engineering.

Virulence and survival mechanisms in *Borrelia* and *Klebsiella* infections – a structural bioinformatics perspective

Mia Åstrand

Biochemistry
Faculty of Science and Engineering
Åbo Akademi University Turku, Finland

2021

From the Faculty of Science and Engineering, Åbo Akademi University, Åbo Akademi Graduate School & National Doctoral Programme in Informational and Structural Biology

Supervised by

Adjunct Professor Tiina A. Salminen

Structural Bioinformatics Laboratory, Biochemistry
Faculty of Science and Engineering,
InFLAMES Flagship Research Center
Åbo Akademi University
Turku, Finland

Reviewed by

Adjunct Professor Salla Ruskamo

Faculty of Biochemistry and Molecular Medicine
Biocenter Oulu
University of Oulu, Finland

and

Adjunct Professor Riikka Ihalin

Department of Life Technologies
Faculty of Technology
University of Turku
Turku, Finland

Opponent

Assistant Professor Ronnie Berntsson

Department of Medical Biochemistry and Biophysics
Umeå University
Umeå, Sweden

ISBN 978-952-12-4075-1 (printed)
ISBN 978-952-12-4076-8 (Digital)
Painosalama, Turku, Finland 2021

Abstract

Bacteria use a wide variety of mechanisms to establish an infection in their host and to compete with other microorganisms for space and nutrients. These virulence and survival mechanisms mainly rely on the functions of proteins, which perform a multitude of highly specialized tasks to ensure bacterial colonization and survival in the host, e.g., through adhesion to host tissues, nutrient import, immune evasion mechanisms and toxin delivery.

As the function of a protein is governed by its structure, determining its three-dimensional structure makes it possible to gain knowledge also about its function. This thesis is focused on virulence factors in *Borrelia* and *Klebsiella* bacteria and I have used structural bioinformatics methods to provide a detailed understanding of how structural features govern the function of a protein.

The BmpA, BmpB, BmpC and BmpD proteins of the Lyme disease-causing bacteria *Borrelia burgdorferi* are an example of how bacteria can survive and reproduce successfully in a host despite a very limited biosynthetic capacity. By analysing 3D structural models of the Bmp proteins, we showed that they are involved in the purine salvage pathway, which is used by the *Borreli*a bacteria to obtain vital purine nucleotides even though they lack the enzymes needed to synthesize them. BmpD was shown experimentally to bind to the purines adenosine and inosine. Our structural comparison showed a high similarity between BmpA, BmpB and BmpD, while BmpC differed significantly from the others, indicating a preference for a different ligand. The evolutionary relationships between the *Borrelia* Bmp proteins were studied in a phylogenetic analysis, which provided an improved classification of the Bmp proteins and revealed significant differences between the Lyme disease-causing and the relapsing fever-causing *Borrelia*.

To successfully colonize a host, *Borrelia* bacteria also need to adhere to host tissues, and this is done through specialized surface-located proteins that bind to receptors on host cells or in the extracellular matrix. Based on our mutational studies, the *Borrelia garinii* DbpA and DbpB proteins are essential for adhesion to cells in the nervous system and mediate adherence by binding to the proteoglycans decorin and biglycan on the surface of HBMECs (human brain microvascular endothelial cells). Negatively charged glycosaminoglycan (GAG) chains on the proteoglycans interact with positively charged lysine residues on the surface of the Dbp proteins and these interactions were inhibited by mutating the lysines. Our structural analysis showed significant changes in surface potential for the mutated proteins, which explained the loss of electrostatic interactions, and highlighted the importance of individual lysine residues for proteoglycan binding.

The highly antibiotic resistant bacterium *Klebsiella pneumoniae* uses secretion systems to inject toxic effectors into host cells or rival microorganisms. The effectors of the type VI secretion system (T6SS) can suppress the host immune system and VgrG proteins play a key role in the function of this system. Our experimental studies showed that *K. pneumoniae* VgrG4 causes a ROS-mediated toxic effect in the host and that its C-terminal DUF2345-containing part is sufficient for ROS induction. Structural modelling predicted that this region has a beta-helical fold, which is characteristic for VgrG spikes. The immunity protein Sel1E prevents the toxic effects of VgrG4 from harming the bacteria itself, and by combining a structural analysis with a study of its evolutionary conservation patterns we highlighted residues possibly involved in interactions with VgrG4.

In summary, this work deepens our understanding of the structure and function of bacterial proteins essential for virulence and survival. Insights into the detailed molecular mechanisms of a protein's function can greatly aid the development of highly targeted novel treatment methods.

Sammanfattning

Bakterier använder sig av en mängd olika mekanismer för att infektera sin värdorganism och för att konkurrera om utrymme och näringsämnen med andra mikroorganismer. Dessa virulens- och överlevnadsmekanismer baserar sig främst på funktionen hos olika proteiner som utför en mängd mycket specialiserade uppgifter för att säkerställa kolonisering och överlevnad i värdorganismen, t.ex. genom adhesion till värdens vävnader, genom upptag av näringsämnen, undvikande av immunförsvaret och leverans av toxiska substanser.

Eftersom ett proteins funktion avgörs av dess struktur kan man få fram information om proteinets funktion genom att bestämma dess tredimensionella struktur. Den här avhandlingen är fokuserad på virulensfaktorer i *Borrelia*- och *Klebsiella*-bakterier och strukturbioinformatikmetoder har använts för att ge en detaljerad förståelse för hur strukturella egenskaper styr ett proteins funktion.

BmpA, BmpB, BmpC och BmpD proteinerna i *Borrelia burgdorferi*, som orsakar borrelios, är exempel på hur bakterier kan överleva och föröka sig framgångsrikt i en värdorganism trots en mycket begränsad biosyntetisk kapacitet. Genom att analysera 3D-strukturmodeller för Bmp proteinerna visade vi att de är involverade i den så kallade purine salvage-rutten, som bakterierna använder för att få tillgång till livsnödvändiga purin-nukleotider, trots att de saknar de enzym som behövs för att syntetisera dem. Experimentella studier visade att BmpD binder till purinerna adenosin och inosin. Vår strukturella jämförelse visade att det finns en stor likhet mellan BmpA, BmpB och BmpD medan BmpC tydligt skiljer sig från de andra, vilket indikerar att den binder till en annorlunda ligand. De evolutionära förhållandena mellan *Borrelias* Bmp proteiner studerades genom en fylogenetisk analys, som resulterade i en förbättrad klassificering av Bmp proteinerna och visade på tydliga skillnader mellan de *Borrelia*-bakterier som orsakar borrelios och de som orsakar återfallsfeber.

För att framgångsrikt kolonisera en värdorganism behöver *Borrelia*-bakterier också fästa sig vid värdens vävnader, och detta görs genom specialiserade ytproteiner som binder till receptorer på värdcellens yta eller i det extracellulära matrixet. Baserat på våra mutationstudier är DbpA och DbpB proteinerna i *Borrelia garinii* väsentliga för adhesion till celler i nervsystemet och förmedlar detta genom att binda till proteoglykanerna decorin och biglykan på ytan av mikrovaskulära endotelceller i människans hjärna (human brain microvascular endothelial cells, HBMECs). Negativt laddade glykosaminoglykankedjor (GAG) på proteoglykanerna interagerar med positivt laddade lysiner på ytan av Dbp proteinerna, och dessa interaktioner inhiberas när lysinerna muteras. Vår strukturella analys visade tydliga skillnader i ytpotential för de muterade proteinerna, vilket förklarar förlusten av de elektrostatiska interaktionerna, och betonar vikten av individuella lysiner för proteoglykanbindning.

Den antibiotikaresistenta bakterien *Klebsiella pneumoniae* använder sig av speciella utsöndringssystem för att överföra toxiska effektormolekyler till värdceller eller konkurrerande mikroorganismer. Effektormolekylerna i utsöndringssystem VI (type six secretion system, T6SS) kan hämma värdorganismens immunförsvar och VgrG proteinerna utgör en viktig del i systemet. Våra experimentella studier visade att VgrG4 från *K. pneumoniae* orsakar en toxisk effekt i värdcellen och att detta styrs genom inducering av oxidativ stress. Proteinets C-terminala DUF2345-innehållande del visade sig vara tillräcklig för detta och genom strukturell modellering visade vi att denna region har en beta-helikal struktur, vilket är typiskt för spike-delen hos VgrG proteiner. Immunitetsproteinet Sel1E förhindrar den toxiska effekten av VgrG4 att skada bakterien själv, och genom att kombinera strukturella analyser med studier av proteinets evolutionära konservationsmönster kunde vi peka ut aminosyror som potentiellt interagerar med VgrG4.

Detta arbete bidrar till en fördjupad kunskap om strukturen och funktionen hos bakteriella proteiner som är nödvändiga för virulens och överlevnad. Insikter i de detaljerade molekylära mekanismerna för ett proteins funktion kan bidra med viktig information vid utvecklingen av nya och specifikt målinriktade behandlingsmetoder.

Table of contents

List of original publications I

Contributions of the author II

Acknowledgements III

Abbreviations V

1. Introduction 1

2. Review of the literature 3

2.1 Bacterial pathogens 3

2.2 Host defense mechanisms 5

2.2.1 Innate immunity 5
2.2.2 Adaptive immunity 6
2.2.3 Antibiotics 7

2.3 Bacterial virulence mechanisms 8

2.3.1 Adhesion 8
2.3.2 Invasion strategies and evasion of host defenses 10
2.3.3 Nutrient acquisition 11

2.4 Interbacterial competition inside the host 14

2.4.1 Contact-independent methods 14
2.4.2 Contact-dependent methods 15
2.4.3 Effectors and immunity proteins 18

2.5 Main virulence mechanisms used by *Borrelia* and *Klebsiella* 19

2.5.1 *Borrelia* 19
2.5.2 *Klebsiella* 21

3. Aims 23

4. Methods 24

4.1 Sequence analysis 24

4.2 Sequence alignment 25

4.3 Modelling protein 3D structures 25

4.4 Model analysis 27

4.5 Phylogenetic analysis 27

4.6 Molecular dynamics simulations 28

4.7 **Visualization and structural analysis** **29**
4.8 **Experimental work** **29**
4.8.1 BmpD crystallization 29

5. **Results and Discussion** **31**
5.1 ***Borrelia burgdorferi*** **– BmpA, BmpB, BmpC and BmpD** **31**
5.1.1 Introduction 31
5.1.2 The *B. burgdorferi* Bmp proteins are substrate-binding proteins 31
5.1.3 Structural characterization of the Bmp ligand-binding site 33
5.1.4 The Bmp proteins are conserved in *Borrelia* species 38
5.1.5 The Bmp proteins play a role in the purine salvage pathway 42
5.2 ***Borrelia garinii*** **– DbpA and DbpB** **44**
5.2.1 Introduction 44
5.2.2 Structural overview of the Dbp proteins 45
5.2.3 Sequence analysis of the *B. garinii* Dbp proteins 46
5.2.4 GAG-binding sites in DbpA and DbpB 46
5.2.5 Protein stability analysis 50
5.3 ***Klebsiella pneumoniae*** **– VgrG4 and Sel1E** **50**
5.3.1 Introduction 50
5.3.2 *K. pneumoniae* VgrGs 51
5.3.3 Structural characterization of VgrG4 52
5.3.4 Sequence analysis and modelling of the immunity protein Sel1E 57
5.3.5 Prediction of potential binding interface between Sel1E and VgrG4 59

6. **Conclusion** **62**

7. **References** **65**

List of original publications

I

Åstrand M, Cuellar J, Hytönen J, Salminen TA (2019). Predicting the ligand-binding properties of *Borrelia burgdorferi* s.s. Bmp proteins in light of the conserved features of related Borrelia proteins. Journal of Theoretical Biology 462: 97–108.

II

Cuellar J, **Åstrand M**, Elovaara H, Pietikäinen A, Guédez G, Salminen TA, Hytönen J. Structural and biomolecular analyses of the BmpD lipoprotein (BB0385) from *Borrelia burgdorferi* reveal a substrate-binding protein of an ABC-type nucleoside transporter. Infect Immun. 2020;88(4):e00962-19.

III

Pietikäinen A, **Åstrand M**, Cuellar J, Glader O, Elovaara H, Rouhiainen M, Salo J, Furihata T, Chiba K, Salminen TA, Hytönen J. Conserved lysine residues in decorin binding proteins of *Borrelia garinii* are critical in adhesion to human brain microvascular endothelial cells. Mol Microbiol. 2021 Jan 29. doi: 10.1111/mmi.14687.

IV

Storey D, McNally A, **Åstrand M**, Sá-Pessoa Graca J Santos J, Rodriguez-Escudero I, Elmore B, Palacios L, Marshall H, Hobley L, Molina Martin M, Cid VJ, Salminen TA, Bengoechea JA. *Klebsiella pneumoniae* type VI secretion system-mediated microbial competition is PhoPQ controlled and reactive oxygen species dependent. PLoS Pathog. 2020;16(3):e1007969.

Contributions of the author

I. The computational work, i.e., sequence alignment, modeling, structural analysis and phylogenetic tree, was done by the author. The author also analyzed all the results and wrote the manuscript.

II. The author did the BmpD crystallization setup and analyzed the final 3D structure of BmpD and its ligand interactions. The corresponding parts of the manuscript were also written by the author.

III. The author performed the computational work, i.e., sequence alignment, modeling and molecular dynamics simulations, analyzed the results and wrote the corresponding parts and all the structural analysis parts of the manuscript.

IV. The author did the VgrG4 and Sel1 models, the structural analyses and the functional site predictions, and wrote these parts in the manuscript.

Acknowledgements

The work for this thesis was carried out at the Structural Bioinformatics Laboratory (SBL), Faculty of Science and Engineering, Åbo Akademi University during 2016 – 2021. It has been a very interesting, rewarding and enjoyable time and I would like to thank everyone who made this possible and everyone who has been a part of these years.

I am very grateful that I got to do my PhD at the SBL, and I would like to thank its leader, Professor Mark Johnson. Most of all, I would like to thank my supervisor Adjunct Professor Tiina A. Salminen, for all your support and patience, and for passing on your extensive knowledge about structural bioinformatics. I have enjoyed this work immensely and I am very grateful that you gave me the chance to work on such interesting projects. I also want to thank Dr. Gabriela Guédez for being my second supervisor and for patiently teaching me about X-ray crystallography.

I would like to thank the National Doctoral Programme in Informational and Structural Biology (ISB) for accepting me and for granting me funding. Thanks especially to the director Professor Mark Johnson, and to the coordinator Fredrik Karlsson for always being so very nice and helpful.

I would like to thank Adjunct Professor Riikka Ihalin and Adjunct Professor Salla Ruskamo for reviewing my thesis and for taking the time to give me valuable comments which helped to improve the thesis. I am also grateful to our collaborators Associate Professor Jukka Hytönen and Professor José A. Bengoechea and their research groups. I especially want to thank my co-author Dr. Julia Cuellar, who I collaborated closely with on two papers and whose positive attitude always made the work feel both fun and easy. I also want to thank Dr. Annukka Pietikäinen, for a very nice and efficient collaboration on my latest publication. I am furthermore grateful to the members of my thesis committee; Associate Professor Jukka Hytönen, Adjunct Professor Outi Salo-Ahen, Adjunct Professor Jarmo Käpylä and Dr. Gabriela Guédez, for their valuable ideas and comments regarding my thesis projects.

I am happy and grateful to have been able to work with all the wonderful people of SBL. The positive, fun and helpful atmosphere has made it a very nice place to work. Leonor Lopes de Carvalho and Käthe Dahlström were the best possible colleagues and friends when I started my PhD-studies and I want to thank you both for all your help, but mostly for all the fun we have had. Leonor, I especially remember all the nice early morning tea breaks we had in the beginning of my PhD. Thanks for sharing all your experience on being a PhD-student. Käthe, you always felt like my big sister in the lab, and it was a pleasure working across from you in our little corner of the office. Mahlet Tamirat, who started out at the same time as me, has been a great support during these years and I am very happy that

I got to do my PhD with you. I was also very glad for your company on the disastrous Berlin conference ☺. Marion Alix and Ida Alanko were my constant companions for lunches and tea breaks during the last few years and I am grateful for all the laughs we have had together. I am also very thankful to Marion, Ida and Mahlet for the virtual coffee breaks during the pandemic. They gave me some much-needed breaks and always put a smile on my face. Polytimi Dimitriou, your sunny smile and laugh brightened up the office and it was always nice to have a talk with you. Parthiban, thanks for your company in the coffee room, and for generously sharing your knowledge about MD-simulations. I am also very grateful to Outi Salo-Ahen, for being part of my thesis committee, for collaborations and for always being so kind and helpful. I also want to extend my thanks to Jukka Lehtonen, for all the help with computer problems and for all your interesting lunch time stories. I also want to thank the people in Peter Slotte's lab, especially Oskar Engberg, for always inviting me for coffee breaks and for teaching me how to work with lipids.

I would also like to thank my family and friends. Tack Mamma, Emma och Malin för att ni funnits där och gett mig en välbehövlig motvikt till jobbet, och för att ni alltid stöttat mig trots att ni inte alltid förstått er på vad jag jobbar med. Leif, tack för du alltid finns där med lugn, tålamod och humor när jag behöver det som mest och för att du hjälpt mig med allt från tekniska problem till diskussioner om grammatik och grafisk utformning. Tack också till Jonna, min bästa vän sedan lågstadiet. Tack för att du alltid finns där trots att vi bor så långt ifrån varandra. Jag är också tacksam för mina ”Åbo-vänner” Jessi, Johanna, Sandra, Anne och Febe. Det är alltid lika roligt att träffa er även om vi nuförtiden är lite mera utspridda. Jag vill också tacka Mia och Hanna, våra bokcirkelträffar, museibesök och spelkvällar har gett mig många roliga och givande andningspauser.

I am also grateful for the generous funding I have received from Magnus Ehrnrooth foundation, Svenska kulturfonden and Tor, Joe och Pentti Borgs Minnesfond, Åbo Akademi.

Abbreviations

3D	Three-dimensional
ABC	ATP-binding cassette
ABPS	Adaptive Poisson-Boltzmann Solver
ATP	Adenosine triphosphate
Bbss	*Borrelia burgdorferi* sensu stricto
Bg	*Borrelia garinii*
BLAST	Basic Local Alignment Search Tool
Bmp	Basic membrane protein
CASP	Critical Assessment of protein Structure Prediction
CDI	Contact-dependent inhibition
Cryo-EM	Cryogenic Electron Microscopy
DAMP	Damage-associated molecular patterns
Dbp	Decorin-binding protein
ExP	Exported protein
GAG	Glycosaminoglycan
I-TASSER	Iterative Threading ASSEmbly Refinement
Kp	*Klebsiella pneumoniae*
LB	Lyme borreliosis
LG	Le and Gascuel
MAPK	Mitogen-activated protein kinase
MD	Molecular Dynamics
ML	Maximum Likelihood
NBD	Nucleotide-binding domain
NCBI	National Center for Biotechnology Information
NF-kB	Nuclear factor-κB
NMR	Nuclear magnetic resonance
OME	Outer membrane exchange systems
Pa	*Pseudomonas aeruginosa*
PAAR	Proline-alanine-alanine-arginine
PAMP	Pathogen-associated molecular patterns
PDB	Protein Data Bank
PnrA	Purine nucleoside receptor A
PRR	Pattern recognition receptor
RF	Relapsing fever
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
ROS	Reactive oxygen species
SBP	Substate-binding protein
SLR	Sel1-like Repeats
TLR	Toll-like receptor
TMD	Transmembrane domain
Tp	*Treponema pallidum*
TPR	Tetratricopeptide repeat
TssX	Type six secretion X
TxSS	Type X secretion system
UniProtKB	Universal Protein Knowledgebase
VgrG	Valine-glycine repeat protein G

1. Introduction

Bacteria are microscopic organisms that thrive in all kinds of environments and have an ability to adapt to the most extreme circumstances. Some bacteria live in our bodies and are vital for the body's normal functions, while other bacteria contain virulence factors that cause disease in their host. The discovery of antibiotics in the 20th century enabled us to treat infections caused by these pathogenic bacteria, but the rapid life cycles of bacteria mean that they are quickly developing resistance to many antibiotics (Hutchings et al., 2019). In combination with our overuse of antibiotics, this has led to an alarming situation where more and more bacteria are rapidly developing antibiotic resistance. At the same time, the rate of antibiotic development has steadily decreased in the last few decades, creating a situation where even common infections can become life-threatening.

Bacteria use complex mechanisms to infect their hosts and to compete with other bacteria for resources. The constant arms race between bacteria and their competitors have led to the development of a multitude of different mechanisms and most bacteria have a large arsenal of weapons for attacking both host cells and rival bacteria (Bliven and Maurelli, 2016). Several mechanisms can often be used simultaneously and are coordinated to produce a specific effect at the right place, at the right time. It has also been suggested that antibiotic resistance could be reduced by using antibiotics that target bacterial virulence mechanisms since this would give a lower selection pressure (Ruer et al., 2015). Most of the virulence mechanisms rely on proteins to perform different functions that harm the target cell in some way, for example by rupturing the target cell or by breaking down its DNA (Webb and Kahler, 2008).

Proteins are large macromolecules that can be considered the workers of a cell. They are involved in all major processes needed to sustain life. On the organism level they play key roles in, for example, the immune system, the metabolism and the nervous system, and on the cellular level they are involved in e.g., DNA replication, enzymatic reactions and transporter functions. Proteins are built from 20 standard amino acids, combined in countless numbers of ways, and with co- and posttranslational modifications, they produce a vast array of proteins of different sizes and shapes. The sequence of amino acids specifies the three-dimensional structure of the protein, which is essential for the function of the protein.

Bioinformatics is a multidisciplinary field that combines biology, mathematics, computer science and statistics to study protein function, evolutionary relationships or predict protein structures (Aamer Mehmood, 2014). The field was originally initiated by Margaret Dayhoff in the 1950s, when she started using computational tools to study protein sequences (Gauthier et al., 2019). To fully understand the function of a protein however, it is vital to know its three-

dimensional structure. This can be determined either through experimental methods, like X-ray crystallography, Cryogenic electron microscopy (cryo-EM), Nuclear magnetic resonance (NMR) or by use of computational tools. Structural bioinformatics is a branch of bioinformatics that uses various computational tools to predict and analyze the 3D structures of proteins (Chandra et al., 2010). From the structure of a protein, its function can be inferred and detailed analyses of how proteins interact with each other and with other molecules can be made. However, when analyzing protein structures, it is important to remember that proteins are highly dynamic molecules that constantly fluctuate between conformations, and in some cases do not form stable structures unless bound to another molecule (Dyson and Wright, 2005; Henzler-Wildman and Kern, 2007).

Since the number of sequenced proteins is much larger than the number of experimentally determined protein structures, the function of many proteins remains unknown. As a result of this discrepancy, novel ways of treating diseases are left undiscovered due to the lack of structural information. Here, computational structure prediction methods can play an important role, as they can predict the structure of a protein much faster and more inexpensively than the experimental methods.

In this thesis, information obtained from protein sequences and structures are combined with the results of functional studies and with analyses of protein interactions to study specific proteins that are involved in host infection and intermicrobial competition. This integrated approach provides detailed information on the proteins and their functional properties and enables us to define their roles in the complex networks of interbacterial and host-pathogen interactions.

2. Review of the literature

2.1 Bacterial pathogens

Bacteria are prokaryotic microorganisms that come in many shapes and sizes and that exist in almost every conceivable environment on earth. Unlike eukaryotes, prokaryotes have a cell wall, which is sometimes surrounded by a sticky capsule, they lack complex organelles and besides their single chromosome they also contain small rings of DNA, called plasmids. Hair-like appendages called fimbriae and pili are used for communication and attachment, and longer appendages called flagella are used for movement. Bacteria are traditionally divided into either Gram-negative or Gram-positive bacteria after Hans Christian Gram, who discovered a staining method which can differentiate bacteria based on their cell wall structure. In Gram-positive bacteria, the cell wall is made up of a plasma membrane, a periplasmic space and a very thick layer of peptidoglycan, containing teichoic and lipoteichoic acids inside it (Cabeen and Jacobs-Wagner, 2005). The Gram-negative cell wall on the other hand is a more multilayered structure, with a lipopolysaccharide (LPS)-rich outer membrane and a periplasmic space between the inner and outer membranes. A thin peptidoglycan layer in the periplasm is connected to the outer membrane via lipoproteins (figure 1).

Many bacteria live in our bodies and form commensal relationships that are mutually beneficial. Our bodies offer a nutrient-rich environment for the bacteria and they in turn provide us with essential vitamins, digest nutrients that our bodies cannot break down, and help prevent colonization by pathogenic bacteria (Brestoff and Artis, 2013). Pathogenic bacteria contain virulence factors that enable them to evade the immune system and cause disease in the host. The virulence of bacteria can be defined as the result of interactions between a microbe and a susceptible host (Casadevall and Pirofski, 2019). Hospital-acquired infections are an example of how normally harmless bacteria can cause disease when the immune system of the host is weakened. In relation to the estimated number of bacterial species, very few species actually cause disease and from an evolutionary perspective it is usually more beneficial for the bacteria to live in a commensal relationship with its host. Despite this, some bacteria do cause disease and it has been speculated that for some bacteria disease is a requirement for successful transmission to a new host.

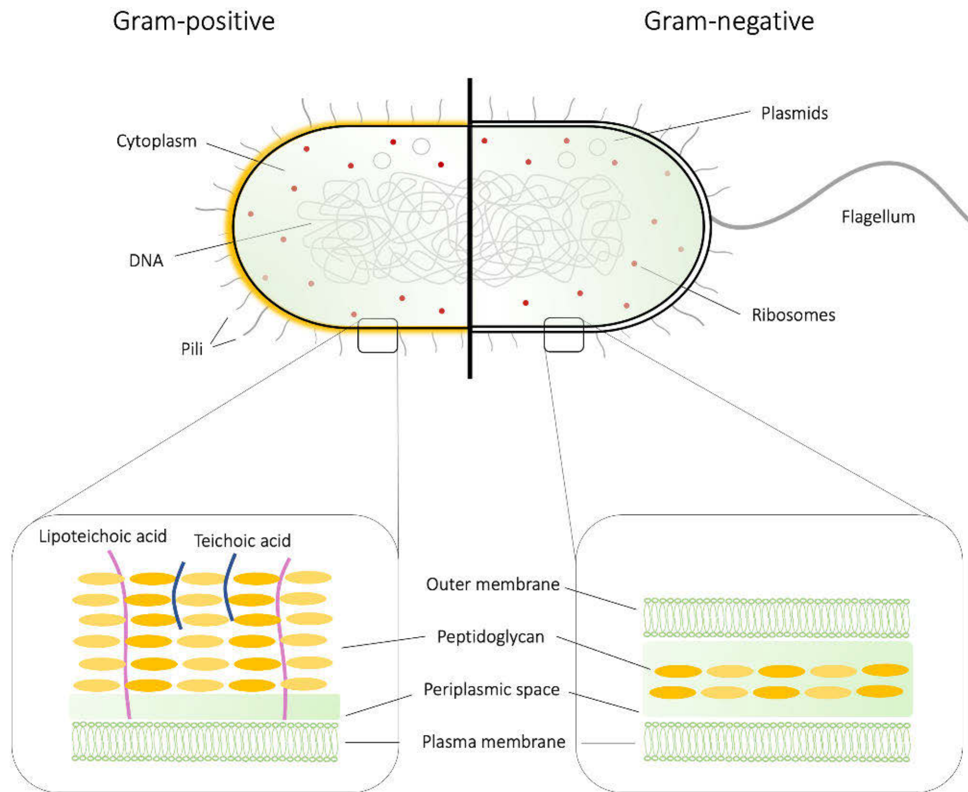


Figure 1. Differences in membrane structure between Gram-positive and Gram-negative bacteria. The figure shows a simplified bacterial cell where the main components are marked. The close-ups show the specific cell wall features of Gram-positive and Gram-negative bacteria. The Gram-positive cell wall consist of a plasma membrane, a thin periplasmic space and a thick peptidoglycan layer with lipoteichoic and teichoic acids. The Gram-negative cell wall consist of two membranes, an inner plasma membrane and an outer membrane. Between them is the periplasmic space which also contains a thin layer of peptidoglycan.

To gain access to a host, bacteria must penetrate its external barriers and then evade its defenses to ensure survival and replication (Sen et al., 2016). For every defense mechanism employed by the host, bacteria have developed effective mechanisms to counteract them, thus creating an arms race between host and pathogen (Webb and Kahler, 2008). Many of these mechanisms are also used in competition with other bacterial species, enabling the bacteria to outcompete their rivals for access to host nutrients.

The aim of this literature review is to give a general overview of bacterial mechanisms used for host infection and interbacterial competition and to emphasize the mechanisms that are relevant for the projects included in this thesis.

2.2 Host defense mechanisms

Pathogenic bacteria entering the body of a host will encounter a wide range of obstacles produced by the defense system of the host. First, physical barriers try to block pathogens from gaining access into the host. When the pathogens have penetrated these initial barriers, they are met by the innate immune system, which produces the first immune responses against the intruder. After a few days, the highly specific adaptive immune response is activated, killing invading pathogens in a highly specific manner. In cases where the host's own defenses are not enough, pharmaceutical antimicrobial compounds (antibiotics) can be used to eradicate the infection.

2.2.1 Innate immunity

Physical barriers, like the skin, the respiratory tract and the gastrointestinal tract, prevent pathogens from entering our bodies, and secretions such as mucus, saliva and tears wash away foreign particles from the epithelial surfaces of the body (Wilson et al., 2002). These barriers also produce chemical substances that are harmful to the pathogens, e.g., antibacterial peptides, enzymes and stomach acid. Antibacterial peptides are produced by many cell types in the body, especially by epithelial cells that come in direct contact with pathogens (Patel and Akhtar, 2017). They are a highly diverse group of molecules that often target the cell membrane or other non-specific targets, thereby making it harder for bacteria to develop resistance (Nguyen et al., 2011). Commensal, non-pathogenic bacteria normally inhabit the epithelial surfaces, and the invading pathogens must compete with them for space and nutrients. Once the pathogen passes these initial barriers, they encounter the cells of the innate immune system, which produces non-specific reactions against invading microbes.

The initial innate cellular response to pathogens is mainly mediated by phagocytic cells like neutrophils and macrophages, which destroy the pathogens by engulfing them in the process of phagocytosis (Nicholson, 2016). The innate immune system cells are activated when specific receptors recognize conserved molecular patterns on the bacteria. These so-called pathogen-associated molecular patterns (PAMPs) are an integral part of bacteria and are essential for their survival and ability to cause infection (Akira et al., 2006). PAMPs are recognized by pattern-recognition receptors (PRRs), which are highly conserved receptors that are continually expressed, both on the inside and the outside of cells (Nicholson, 2016). The Toll-like receptor (TLR) family are among the most important pattern recognition receptors and can detect a large variety of microbial patterns (Akira and Takeda, 2004). The immune system can also be activated by damage-associated molecular patterns (DAMPs), which are molecules released by the body as a result of the tissue damage and cell lysis that occurs during the infection (Kaur and Secord, 2019). In addition to these

molecular patterns, the body can react to the loss of specific molecules that are expressed by healthy and normal cells.

Induction of oxidative stress is another way for the host to combat pathogens (Pohanka, 2013). Reactive oxygen species (ROS) are normally produced as a result of aerobic respiration and in the presence of free metals but can also be induced as a defense mechanism against pathogens. ROS damages macromolecules and disrupts signaling pathways by causing strand breaks in DNA or by oxidating proteins, lipids or DNA (Auten and Davis, 2009). During infections, ROS is triggered by the phagocytosis of bacteria and a so-called respiratory burst is activated inside the phagosome. ROS can also be released into the extracellular space, targeting bacteria that escapes phagocytosis (Nguyen et al., 2017).

The complement system plays a major role in the initial immune response against bacterial infections and consist of a network of proteins that together enhances the ability of the immune system to destroy pathogens, by marking the pathogens for phagocytosis, by producing proinflammatory molecules or by inducing cell lysis (Dunkelberger and Song, 2010). Together, these initial responses of the innate immune system activate the highly specific reactions of the adaptive immune system (Iwasaki and Medzhitov, 2015).

2.2.1.1 Nutritional immunity

Another defensive strategy against invading pathogens is the sequestration of vital nutrients. Many enzymes and proteins require trace minerals like iron, zinc and manganese to function and these elements are consequently vital for both host and pathogen. The methods employed by the host for preventing pathogen access to minerals is called nutritional immunity. Sequestration of iron is the most common form of nutritional immunity and most of the iron inside vertebrate bodies is stored intracellularly, either in the iron storage protein ferritin or in the heme of hemoglobin or myoglobin (Skaar, 2010). In response to inflammatory signals the body can prevent iron transfer into the circulation by inhibiting ferroportin, an iron exporter, and by secreting lactoferrin, which binds to Fe^{3+} with high affinity, thereby further restricting the pathogen's access to iron (Núñez et al., 2018). Any remaining extracellular free iron is quickly bound to transferrin ensuring that virtually no free iron is accessible to pathogens (Cassat and Skaar, 2013). However, as the hosts have developed systems to sequester minerals from the pathogens, the pathogens have consequently developed mechanisms for circumventing them (see more details in section 2.3.3.2) (Hennigar and McClung, 2014).

2.2.2 Adaptive immunity

Compared to the innate immune system, it normally takes several days before the adaptive immune system is activated, but then it provides a highly specific

response to pathogens and also creates an immunological memory. The adaptive immune response relies on two main types of lymphocytes, B cells and T cells. Both cell types have surface receptors that bind with high specificity to molecules recognized as non-self, antigens. B cells recognize antigens on the surface of pathogens whereas T cells only recognize fragments of antigens exposed on the surface of antigen-presenting cells (Actor, 2019). Fragments of antigens can be presented on the surface of many cell types, for example on phagocytic cells of the innate immune system, or on the surface of B cells.

When B cells recognize a specific antigen, they are triggered to start dividing and become antibody-producing plasma cells or memory cells (Bonilla and Oettgen, 2010). Binding of an antibody to an antigen activates, for example, the classical pathway in the complement system, resulting in cell lysis, or the phagocytic cells, increasing phagocytosis (Forthal, 2014). It can also cause cells to aggregate or be immobilized, or the antibodies can physically prevent pathogens from attaching to host cells. When a T cell has been activated, it proliferates into subtypes that perform different functions: T killer cells destroy infected cells whereas T helper cells secrete cytokines, which activate macrophages or stimulate B cells to increase the production of antibodies.

2.2.3 Antibiotics

Although not part of the natural defense systems of the host, antimicrobial compounds in the form of pharmaceuticals, i.e., antibiotics, can also provide a significant defense against infectious bacteria. Antibiotics are antimicrobial compounds designed to kill or impede the growth of pathogenic bacteria in the body (Hutchings et al., 2019). Many antibiotics originate from compounds produced naturally by plants, bacteria or fungi, who use them as chemical weapons or as defense mechanisms against competitors. Antibiotics mainly target essential functions or structures of the bacteria and are generally classified according to their mechanism of action (Kapoor et al., 2017). Beta-lactams and glycopeptide antibiotics prevent cell wall synthesis by disrupting the formation of the peptidoglycan layer (Bush, 2012). Macrolides and tetracyclines are examples of antibiotics that target the protein synthesis machinery. They bind to ribosomal subunits and inhibit protein synthesis by preventing tRNA binding or by causing unfinished protein chains to detach from the ribosome (Mccoy et al., 2011). DNA synthesis can be disrupted by fluoroquinolones, which inhibit the DNA topoisomerases needed for DNA replication (Aldred et al., 2014). Vital metabolic pathways can also be targeted, e.g., the synthesis of folic acid, needed for DNA synthesis, is inhibited by sulfonamides and trimethoprim (Fernández-Villa et al., 2019).

2.3 Bacterial virulence mechanisms

Bacteria have developed a wide range of mechanisms to subvert the host immune system and to sustain an infection (Wilson et al., 2002). They use adhesion mechanisms to attach themselves to a host cell, produce toxins or effectors that destroy or damage host cells and evade the host defense mechanisms by masking themselves in different ways to prevent recognition by the host.

2.3.1 Adhesion

In order to colonize a host, the bacteria first need to attach themselves firmly to the host surfaces. Mechanical forces (coughing, saliva, blood flow etc.) in the host constantly try to remove bacteria from its surfaces, and forming a stable adherence is therefore a crucial first step in the infection process (Wilson et al., 2002). When bacteria come into contact with the host surfaces, they can sense changes in their environment, which in turn trigger changes in metabolism, respiration and expression of virulence-specific genes (Stones and Krachler, 2016). There are a large number of adhesion molecules (adhesins) on the surface of bacteria, and these can form a wide variety of interactions with the host surfaces. Some interactions are transient and nonspecific whereas other interactions are stable and highly specific. The different kinds of adhesins are normally expressed in a coordinated fashion at different stages of the infection (Klemm and Schembri, 2000).

Adhesins are divided into two main groups: fimbrial (pili) and afimbrial adhesins (Pizarro-Cerdá and Cossart, 2006; Ribet and Cossart, 2015; Wilson et al., 2002). Fimbriae are short hair-like polymeric structures on the surface of bacteria with a protein at the tip, which determines the binding specificity to molecules on the host cell. Afimbrial adhesins are bacterial surface proteins that bind to a variety of molecules on the host surface, like integrins and cadherins, as well as components of the extracellular matrix (collagens, laminins, proteoglycans etc.) and usually mediate more intimate, short-range contacts than the fimbria (Wilson et al., 2002). The adhesin OspC (Outer Surface Protein C) is an example of an afimbrial adhesin and plays a crucial role in *Borrelia* infections by binding to the extracellular matrix components dermatan sulfate and/or fibronectin. Bacteria with OspC that can bind to dermatan sulfate are capable of colonizing mice joints, whereas bacteria with OspC that can only bind to fibronectin are unable to colonize joints (Lin et al., 2020).

Adhesion is also an important first step in the formation of biofilms. Bacteria are capable of aggregating into multicellular communities with other microorganisms and together they secrete a polysaccharide-rich matrix that can protect them against many host defense mechanisms, including antibiotics (Ribet and Cossart, 2015).

2.3.1.1 Bacterial adhesion to proteoglycans

Many pathogens adhere to host tissues by binding to proteoglycans, which are widely expressed in many tissues. Proteoglycans are glycosylated proteins found on host cell surfaces, intracellular compartments and in the extracellular matrix, where they mediate cellular processes like adhesion, signaling and motility (Rostand and Esko, 1997). They consist of a core protein that is covalently bound to one or more glycosaminoglycan (GAG) chains (figure 2) (Varki et al., 1999), which are made up of repeating disaccharide units consisting of an amino sugar and a uronic acid. Each GAG is made up of a specific combination of disaccharide units and negatively charged sulfate groups are attached to some of them, creating a unique sulfation pattern. These sulfate groups mediate electrostatic interactions to GAG-binding proteins (Hileman et al., 1998). Proteoglycans are thus highly variable molecules and their distribution and composition can vary depending on tissue type (García et al., 2016). Which tissues a pathogen can colonize may be determined by its preference for a specific GAG. *B. burgdorferi* has, for example, been shown to bind to proteoglycans in various host tissues (e.g. endothelial and neural tissues) and to bind to different types of GAGs depending on the tissue (Leong et al., 1998).

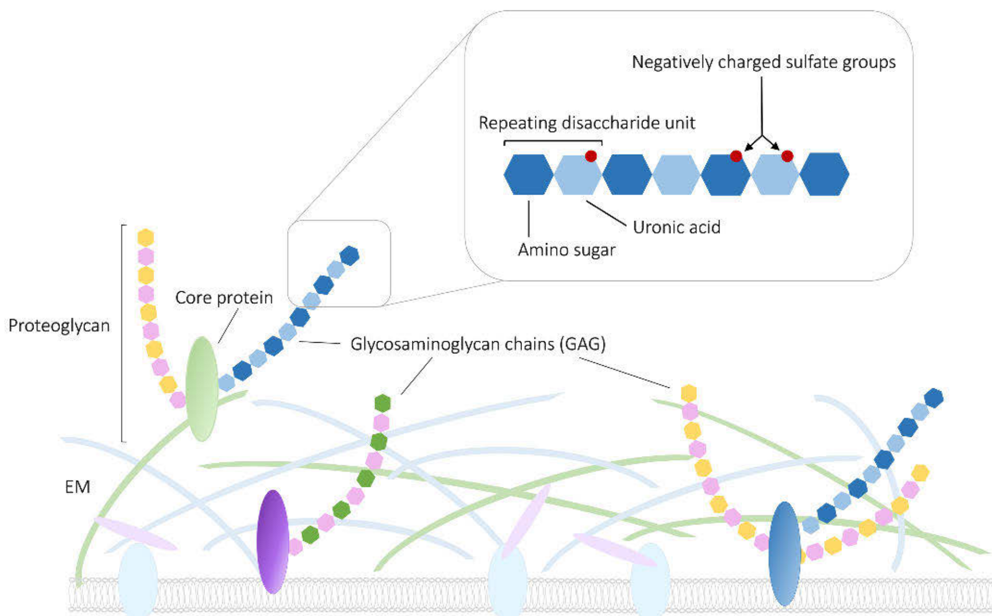


Figure 2. General structure of proteoglycans and glycosaminoglycans (GAGs) shown as components of the extracellular matrix (EM). Proteoglycans consist of a core protein and attached GAG chains. A GAG chain is made up of repeating disaccharide units consisting of an amino sugar and a uronic acid, which can vary between different GAGs. Negatively charged sulfate groups are attached to the GAG chains in specific sulfation patterns. The GAG chains interact with other molecules by forming electrostatic interactions through the sulfate groups.

2.3.2 Invasion strategies and evasion of host defenses

After adhering to a host cell, many bacteria need to gain further access into the host. Most bacteria have either an extracellular or an intracellular lifestyle (Wilson et al., 2002), but some bacteria can use both strategies. Extracellular bacteria produce enzymes that break down the tissue barriers and then establish a niche within the tissue, but without entering host cells. Intracellular pathogens on the other hand penetrate the host cells and modify the cells internal environment in order to survive within the host cell without being degraded (Thakur et al., 2019).

Avoiding the effects of antimicrobial peptides is one vital strategy used by bacteria to evade the immune system (Reddick and Alto, 2014; Sperandio et al., 2015). This can be done e.g. by changing the charge of membrane lipids to function as an electrostatic buffer (Ernst et al., 2009), to secrete proteases that break down antimicrobial peptides (McGillivray et al., 2009) or to use transporters to export the antimicrobial peptides or import them for destruction (Eswarappa et al., 2008; Shafer et al., 1998). Another evasion mechanism is to mask the bacterial surface in a capsule consisting of polysaccharides, which hides the specific bacterial molecules recognized by the immune system, while still allowing adhesins to penetrate (Finlay and McFadden, 2006; Paton and Trappetti, 2019). Lipopolysaccharide, flagella and peptidoglycan are some of the most easily recognizable bacterial surface molecules and bacteria use many different methods to hide these from the immune system (Hajam et al., 2017; Kawasaki et al., 2004; Sorbara and Philpott, 2011). By making small alterations to these kinds of molecules, the bacteria can avoid being recognized by the immune system. Antigenic variation is an example of how bacteria can modify the surface antigens recognized by the adaptive immune system, and thereby avoid immune activation (Palmer et al., 2016). As many immune cells are phagocytic, bacteria have developed mechanisms for avoiding phagocytosis or intracellular killing in phagolysosomes, e.g. by causing lysis of the phagocytic cell or by preventing opsonization (Uribe-Querol and Rosales, 2017). Molecular mimicry is a method used by bacteria to hide themselves from the host immune system by mimicking host cell surface molecules (Brown et al., 2008). Human cells are coated in sialic acid residues and the bacteria *Neisseria meningitidis* can escape detection by decorating its surface with these residues (Parsons et al., 1996). Bacteria can also modify their PAMPs in such a way that the PRRs can no longer recognize them (Matsuura, 2013).

The pathogens have also evolved mechanisms to avoid the damages caused by oxidative stress, for example by repairing damaged DNA and proteins, or by producing antioxidants and detoxification enzymes that neutralize ROS (Reniere, 2018). Furthermore, it has also been shown that some bacteria can induce ROS in target cells and use it to their own advantage (Dong et al., 2015; Hersch et al., 2020). Bacteria can also use other methods for causing host damage, e.g., by

attacking crucial signaling pathways (e.g. MAPK and NF-kB) or protein secretory pathways (Duesbery et al., 1998; Kim et al., 2007; Sanada et al., 2012). Many of the specific effectors used to attack host cells are also used in interbacterial competitions and will be discussed in more detail in section 2.4.

2.3.2.1 Antibiotic resistance

Bacteria are constantly exposed to antimicrobial compounds in their natural environments and develop resistance to the antibiotics through the process of natural selection. Resistance can be acquired through mutations or through horizontal gene transfer (Holmes et al., 2016). In horizontal gene transfers, the bacteria obtain genetic material from the environment or by direct transfer from other bacteria. This mechanism enables rapid acquisition of antibiotic resistance genes and is the main cause of the increasing antibiotic resistance found among pathogenic bacteria today (Munita and Arias, 2016).

The mechanisms used to confer antibiotic resistance can be categorized based on the biochemical pathways used (Munita and Arias, 2016). A commonly used strategy is to chemically alter or destroy the antibiotic compound. The antibiotic can be modified e.g., by acetylation or phosphorylation, and often results in a steric hindrance that prevents the antibiotic from binding to its target. Bacteria can also prevent the antibiotic from reaching its target by decreasing its uptake into the cell or by actively pumping the antibiotic out of the cell through efflux pumps. The target site can also be protected in different ways to prevent binding of the antibiotic, e.g., the tetracycline resistance determinant Tet(M) can remove tetracycline from the ribosome and then alter the conformation of the ribosome to prevent tetracycline binding (Dönhöfer et al., 2012). Modification of the target site is another way to prevent antibiotic binding. This can be done e.g., through point mutations or by chemical modifications such as methylation.

2.3.3 Nutrient acquisition

Nutrients, like carbohydrates, nucleic acids, amino acids, lipids and transition metals like iron, zinc and manganese, are needed as building blocks for cells and to produce energy for cellular growth and replication. The metabolic requirements of bacteria can vary greatly, and most bacteria have highly flexible metabolic systems that can be quickly adapted to differing environments. Pathogenic strains generally have a higher degree of flexibility than non-pathogenic strains (Passalacqua et al., 2016). The host environment contains plenty of nutrients and potential niches for bacteria. However, the host metabolism is also tightly regulated and in response to pathogens it can be even further restricted as a means to prevent bacterial colonization. In order to gain a foothold, the bacteria use several different strategies to overcome these defenses.

Inside the host, many of the nutrients needed by bacteria are locked inside large macromolecules and are not easily accessible to the bacteria. However, the essential carbon and nitrogen compounds found in the macromolecules can be obtained through the action of degradative enzymes like proteases and phospholipases (Lehman et al., 2019). For the intracellular bacteria *Mycobacterium tuberculosis*, host cell lipids are the main carbon and energy source (Rameshwaram et al., 2018) and they use lipolytic enzymes to hydrolyze host lipids into free fatty acids that can be used for the bacteria's own needs. Some bacteria can hijack host pathways and use them for their own purposes. The intracellular bacteria *Legionella pneumophila* ubiquitinates host cell proteins, which target them for degradation, and thereby produces free amino acids for its own use (Price et al., 2011). Genes involved in metabolic pathways can also be acquired through horizontal gene transfer, just like other virulence genes (Abu Kwaik and Bumann, 2015), and enable the bacteria to take advantage of host nutrients by providing novel ways of accessing host nutrients.

2.3.3.1 Nucleic acid uptake

Nucleotides play essential roles as building blocks for RNA and DNA, provide energy for cellular processes (adenosine triphosphate (ATP)), function as cofactors and play key roles in signaling and regulation (Kilstrup et al., 2005). Most bacteria are capable of producing nucleotides *de novo* (figure 3), whereas others must obtain nucleobases and nucleosides from the surrounding environment and use salvage pathways to convert them into nucleotides. Among others, the lactobacilli (Kilstrup et al., 2005) and the spirochetes (Pettersson et al., 2007) are unable to *de novo* synthesize nucleotides and rely on salvage pathways. The Lyme disease-causing *Borrelia* spirochetes have a parasitizing lifestyle and need to obtain most of its nutrients from the host (Radolf et al., 2012). Some of the key enzymes of the salvage pathway are missing in these bacteria, and they are therefore, for example, completely dependent on access to host deoxynucleotides (Lawrence et al., 2009).

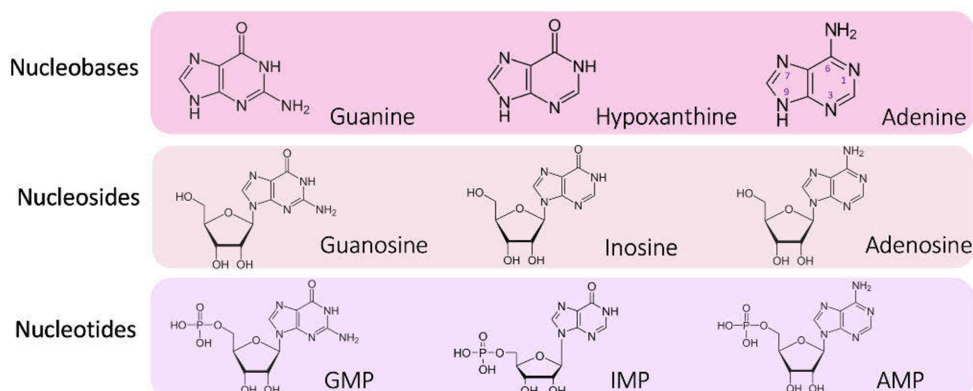


Figure 3. Chemical structures of nucleobases, nucleosides and nucleotides (nucleoside monophosphates). The conventional numbering of nucleobase atoms is shown in the adenine structure. Figure adapted from Publication II.

2.3.3.2 Metal uptake

Iron, zinc, copper and manganese are crucial for a wide range of cellular processes, such as amino acid synthesis, DNA synthesis, as enzyme cofactors, and in electron transport, and are thus essential compounds for most living organisms. However, at higher levels these metals are toxic to cells, and are therefore tightly regulated (Passalacqua et al., 2016). The host iron levels are always kept at a level too low for optimal bacterial growth and bacteria therefore require specialized uptake systems to obtain the needed iron (Rohmer et al., 2011). Many bacteria also use the low iron levels as a signal that they are inside a vertebrate host and need to activate virulence factors. Bacteria produce high-affinity siderophores that can bind to free iron, and the siderophore-iron complexes are then taken up by the bacteria (Hennigar and McClung, 2014). The host counteracts this by producing siderocalins, which bind to the siderophores and prevent their uptake by the bacteria. Some bacteria, like *Neisseria*, have no siderophores but can hijack the siderophores of other bacteria (Cornelissen, 2018). Iron can also be accessed from hemoglobin, using heme-binding proteins. The Isd (iron-responsive surface determinant) system in *Staphylococcus aureus* consists of an efficient chain of heme-binding proteins that captures heme directly from hemoglobin and transports it into the cell (Grigg et al., 2010). Some bacteria have taken a different approach to the problem and eliminated the need for iron altogether. The genomes of Lyme disease *Borrelia* encode very few metal-requiring proteins and has substituted iron for manganese in these, enabling the bacteria to survive in iron-limited environments without specialized iron-acquisition systems (Posey and Gherardini, 2000).

2.3.3.3 ABC-transporters in nutrient import

ABC (ATP-binding cassette) transporters are used as importers for a wide range of molecules (amino acids, metal ions, vitamins, peptides etc.) and have been shown to be essential for bacterial survival inside the host (Tanaka et al., 2018). ABC transporters are found in both prokaryotes and eukaryotes and is one of the largest families of transporter systems (Davidson et al., 2008a). They use the hydrolysis of ATP to transport substrates over cell membranes (Rees et al., 2009) and consist of two transmembrane domains (TMD) and two nucleoside-binding domains (NBD) (figure 4). The TMDs form a translocation channel through the membrane and the NBDs bind to and hydrolyze ATP, powering the transport of substrates through the ABC-transporter. In bacteria, the ABC-transporters require a substrate-binding protein (SBP) to deliver the substrate to the membrane-bound transporter complex (Licht and Schneider, 2011).

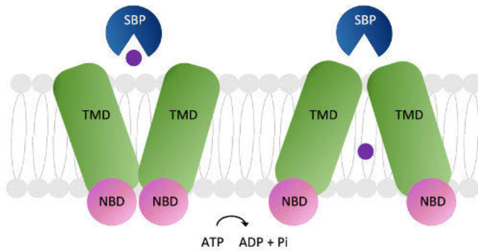


Figure 4. Bacterial ABC-transporter systems consist of two transmembrane domains (TMDs), two nucleotide-binding domains (NBDs) and a substrate-binding protein (SBP). The general function of ABC-transporter systems is shown. First, the SBP binds to a substrate molecule (purple) and transports it to the membrane bound TMDs. The TMDs open in an ATP-dependent manner and allow the substrate molecule to pass through the membrane. Figure adapted from Publication II.

2.4 Interbacterial competition inside the host

Inside a host, bacteria also need to compete with other bacteria, pathogenic or commensal, for nutrients and space (Stubbendieck and Straight, 2016). Bacteria can out-compete other species by depleting all available nutrients (exploitative competition) or in a more direct way by producing toxins that inhibits the growth or division of other cells or by physically damaging the cell (interference competition)(García-Bayona and Comstock, 2018). Bacteria employ a large variety of antibacterial toxins, ranging from small molecules to large proteins, and some toxins target only closely related strains while others target a wide variety of strains (Hibbing et al., 2010). Small molecules can normally pass through the target cell membrane by simple diffusion, whereas larger proteinaceous toxins require more advanced transport mechanisms. The methods used for interbacterial competition can be broadly divided into contact-independent and contact-dependent methods. Studies have shown that bacteria can respond to threats in a distance-dependent way, where contact-independent methods are used to respond to threats that are further away, whereas contact-dependent methods are used for more immediate threats (Westhoff et al., 2017).

2.4.1 Contact-independent methods

Contact-independent methods do not require physical contact between cells and are mainly mediated through secretion of toxic substances. Bacteria have a large arsenal of toxic molecules that are released from the cell through membrane pores, through active transport or by means of membrane vesicles (Granato et al., 2019) and some larger protein toxins are even released through cell lysis (García-Bayona and Comstock, 2018). These molecules are highly diverse and their functions are still largely unknown, however, many of them affect the growth and development of the target bacteria (Stubbendieck and Straight, 2016).

Many of the mechanisms used by bacteria rely on chemical toxins, molecules that chemically interfere with the function of another cell (Granato et al., 2019). Bacteriocins are a large and diverse group of antibacterial peptides or proteins, which normally target closely related bacterial strains, and that require specific receptors on the target cell (Hassan et al., 2012). Some bacteria use biological weapons in the form of phages, viruses incorporated into the bacterial genome, and these can be released to kill other bacterial cells by injecting their own genetic material into the cell (Patz et al., 2019). Mechanical weapons can be used to physically damage the target cell, e.g., tailocins, which are phage-like molecules that puncture the membrane of a target cell (Granato et al., 2019). Bacteria can also release extracellular vesicles, both to defend against attacks from other bacteria and to deliver toxic molecules to competing bacteria (Stubbendieck and Straight, 2016). The vesicles bud off from the surface of the producing bacteria and fuse with the membrane of the target bacteria.

2.4.2 Contact-dependent methods

The contact-dependent methods require close contact between the bacteria and its target and rely on a transfer of proteins directly into a target cell or into the extracellular environment (Green and Mecsas, 2015). The secreted proteins can be used to induce toxic effects in the host, to facilitate adhesion, to obtain nutrients from the environment or to compete with other bacteria in establishing a niche. There are two main types of mechanisms used: contact-dependent growth inhibition (CDI) systems and secretion systems.

In the CDI system, an outer membrane beta-barrel transporter (CdiB) transports CdiA, a large protein containing a toxin domain, to the cell surface where it binds to a specific receptor on the target cell (Aoki et al., 2005; Jones et al., 2017; Willett et al., 2015). An immunity protein (CdiI) also belongs to this system and its role is to prevent autointoxication. The toxin domain of CdiA can be highly variable and enable the bacteria to deliver toxins with a wide range of activities. However, most of the toxins recognized so far have enzymatic activities that target nucleic acids (Allen and Hauser, 2019).

Specialized secretions systems are used by bacteria to secrete proteins (effector molecules) across membranes (Green and Mecsas, 2016; Klein et al., 2020) and these systems are also used by pathogenic bacteria to transport virulence factors into host cells (figure 5). Some secreted proteins are toxic to the host cell or disrupt vital functions, whereas other proteins help the bacteria to attach to cells, establish themselves in a specific niche or to compete with other bacteria in the environment. Bacteria can transfer a large number of different effector molecules by using the secretion systems and since some effector molecules can have synergistic or conflicting effects the translocation have to be strictly coordinated and timed in order to produce the right effects at the right time (Stones and Krachler, 2016).

In both Gram-negative and Gram-positive bacteria, the Sec (general secretion) and Tat (twin arginine translocation) pathways are used for general transport of proteins between intracellular compartments and over the cytoplasmic membrane (Frain et al., 2019; Tsirigotaki et al., 2017). These secretion systems are highly conserved within all domains of life and proteins transported by these systems normally remain inside the cell or in the periplasmic space. However, they work in concert with secretion systems that span only the outer membrane and deliver effectors to these systems for further transport outside the cell.

Since Gram-negative bacteria consist of an inner and an outer membrane separated by the periplasmic space, secretion of proteins to the outside of the bacteria requires specialized systems that can penetrate these layers. To date there are nine known secretion systems, all of them secreting different kinds of substrates using different mechanisms (Bhoite et al., 2020; Green and Mecsas, 2016; Lasica et al., 2017). T2SS (Type 2 secretion system), T5SS and T8SS can only secrete proteins across the outer membrane and relies on the Sec and Tat pathways to first deliver the proteins into the periplasmic space (Fan et al., 2016; Korotkov and Sandkvist, 2019). The T8SS is dedicated exclusively to the transport of curli as part of the curli biogenesis pathway (Bhoite et al., 2020). T1SS, T3SS, T4SS and T6SS are capable of transporting proteins across both the inner and outer membranes and, except for T1SS, these systems can also penetrate the host membrane and transport proteins directly into the host cell (Cherrak et al., 2019; Dey et al., 2019; Sgro et al., 2019; Spitz et al., 2019). Recently it was demonstrated that a *Yersinia pseudotuberculosis* T6SS system can also release effectors directly into the extracellular space in a contact-independent manner (Song et al., 2021). The effectors are then taken up by the host cell through specific proteins on the outer membrane. The T9SS has been discovered only in *Bacteroidetes*, where the system seems to be involved either in motility or function as a weapon (Lasica et al., 2017). Its major components have been identified but its mechanism of action is still not fully known.

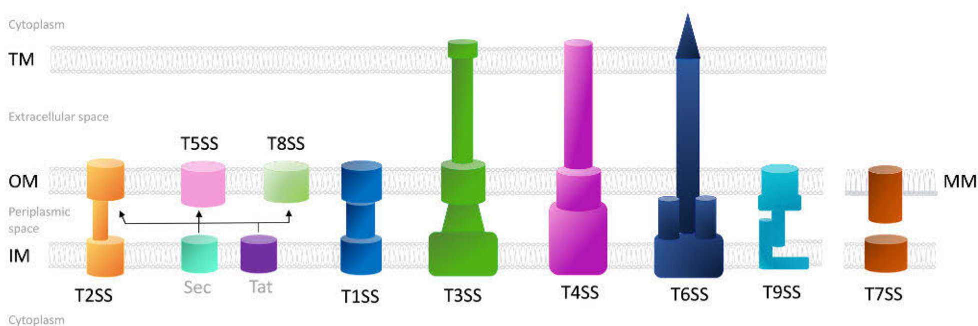


Figure 5. Simplified overview of currently known bacterial secretion systems. T2SS, T5SS and T8SS require the help of the Sec and Tat secretion systems to first transport substrates into the periplasmic space. T3SS, T4SS and T6SS can transport their substrates directly into a target cell. TM: target cell membrane, OM: Outer membrane, IM: Inner membrane, MM: mycomembrane.

Gram-positive bacteria lack an outer membrane but have a very thick cell wall consisting of peptidoglycan. In addition to the Sec and Tat pathways, many Gram-positive bacteria use an additional factor (SecA2) for Sec secretion (Tsirigotaki et al., 2017). Proteins can also be released into the extracellular space through passive diffusion through the peptidoglycan layer. Certain Gram-positive bacteria, like the Mycobacteria, contain a so called mycomembrane, a dense, hydrophobic lipid layer, outside the cell wall and a T7SS system have been identified in these bacteria and is thought to transport proteins across both the inner membrane and the mycomembrane (Houben et al., 2014).

2.4.2.1 T6SS

The T6SS was first described in *Vibrio cholerae* as a novel mechanism for extracellular protein secretion (Pukatzki et al., 2006) and was shown to be active during chronic *Pseudomonas aeruginosa* infections (Mougous et al., 2006). Bacteria use the T6SS both to infect its host and to compete with other bacteria (Hood et al., 2010; Murdoch et al., 2011; Schwarz et al., 2010). It is a large complex of proteins that together form a sophisticated machinery that can transfer effector molecules into target cells (figure 6). The complex is structurally similar to the bacteriophage injection systems (Leiman et al., 2009) and is encoded in a gene cluster with 13 conserved core genes, which are all essential for the function of the system (Cianfanelli et al., 2016). In addition to the core genes, the clusters may encode accessory proteins and effectors and other proteins that may be needed for the construction or regulation of the system (Cianfanelli et al., 2016).

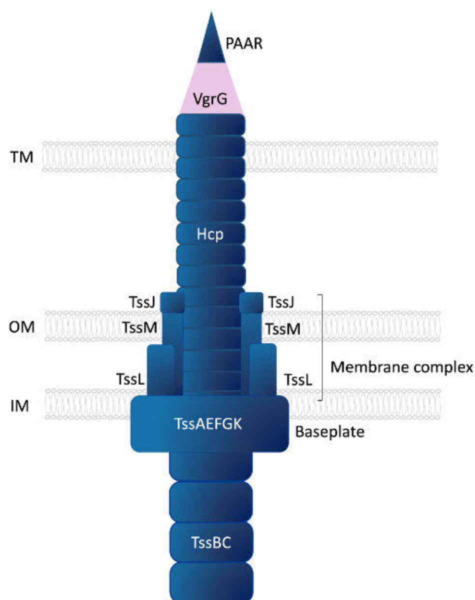


Figure 6. The T6SS system. The protein complex consists of a baseplate structure (TssAEFGK), a membrane complex (TssLMJ), an inner tube consisting of stacked rings of Hcp, an outer tube made up of TssBC, and VgrG and PAAR proteins at the tip of the structure. The membrane complex spans both the inner (IM) and outer membranes (OM) and the tip of the structure penetrates the target cell membrane (TM).

The secretion machinery is anchored to the bacterial membrane through a membrane complex, which consists of the proteins TssJ (Type six secretion), TssL and TssM. A baseplate structure is connected to the membrane complex and provides a platform for assembly of the tube and sheath structures. The baseplate is formed by five different proteins: TssA, TssE, TssF, TssG and TssK. An inner tube is made up of stacked rings of Hcp, which show a structural similarity to gp19 of the bacteriophage T4 tail-tube (Leiman et al., 2009). Outside of the inner tube, the proteins TssB and TssC form a contractile sheath, which can be contracted and thereby cause the inner tube to be pushed through the cell wall and into the target cell. The inner tube is capped by a VgrG protein and a PAAR repeat protein, which form a sharp tip that penetrates the cell membranes and are released into the target cell when the complex pierces the cell wall. VgrG is structurally very similar to the gp27/gp5 complex that forms the spike of the bacteriophage T4 (Leiman et al., 2009). There can be more than one type of VgrG protein in a species and these can have slightly different functions and be associated with different effectors and PAAR proteins (Cianfanelli et al. 2016b). In *Serratia marcescens*, cargo effectors show a preference for a certain VgrG proteins and specific PAAR-VgrG combinations are required for proper assembly of the T6SS (Cianfanelli et al. 2016b).

2.4.3 Effectors and immunity proteins

Bacteria use a huge array of different toxins and effectors to attack neighboring bacterial cells, and many of them are also used to attack host cells. The word effector refers to molecules that require specialized transporter systems for their delivery and exert their function in concert with other effectors in a coordinated fashion. Toxins and effectors often perform similar functions, but toxins do not require any specialized transporter complexes and they function independently of other toxins (Galán, 2009). Effectors are highly diverse in size, shape and sequence but they still have some characteristics in common (Ruhe et al., 2020). The N-terminal part is normally a conserved region that mediates export and delivery into the target cell, whereas the C-terminal part contains the toxic activity. The effectors are also modular and carry different kinds of toxin domains, which can be broadly divided into groups based on the kind of mechanisms they use (Ruhe et al., 2020; Russell et al., 2014).

Phospholipases and pore-forming toxins both target the cell membrane (Russell et al., 2014). Phospholipases are primarily delivered by T6SSs and are often fused to Hcp, VgrG or PAAR proteins. They hydrolyze the cell membrane phospholipids and disrupt the cell membrane. Pore-forming toxins change the permeability of the bacterial membranes, causing a wide range of effects on the target cell, from activation of inflammasomes and disrupted protein synthesis to cell death (Peraro and Van Der Goot, 2016). The cell wall peptidoglycan is another target for effectors. Effectors can disrupt it in two ways: by preventing the synthesis of the peptidoglycan or by enzymatically degrading it (Ruhe et al.,

2020). Both methods ultimately cause lysis of the target cell. Several enzymatic effectors, e.g. DNases, deaminases and NADases, are targeted to the cytoplasm (Jurėnas and Journet, 2020), where they cause damage to DNA or to other essential macromolecules in the cell. Extracellular metallophore effectors, which scavenge for zinc, copper or manganese and import them back into the bacteria, have also been discovered (Han et al., 2019; Si et al., 2017; Wang et al., 2015).

Each effector is normally encoded with an immunity protein that protects the producing bacteria from the toxic effects of its own effectors and from those of sibling bacteria. The immunity proteins often bind to the active site of the effector to physically block the site or some can even reverse the effects caused by the effector (Coulthurst, 2019; Ting et al., 2018). Immunity genes have also been found without a cognate effector, so-called "orphaned" immunity genes, which provide the bacterium with protection even if it lacks its own effector (Hersch et al., 2020). They are often genes acquired through horizontal gene transfer or immunity genes whose effector have been lost, and they can protect the bacterium against effectors from other strains (Coulthurst, 2019). Immunity proteins can either be expressed continuously or be induced by the presence of other bacteria, by nutrient availability or by the growth phase of the bacteria (Hersch et al., 2020).

2.5 Main virulence mechanisms used by *Borrelia* and *Klebsiella*

All bacteria have their own unique ways of surviving and reproducing inside a host, shaped by the evolutionary arms race that takes place between each pathogen and its host (Bliven and Maurelli, 2016). The pathogen has developed strategies that enable it to survive in its specific host environment, and these are continuously refined in response to new defense mechanisms by the host. As the host environment changes, bacterial genes that are no longer necessary may be discarded and new genes that provide an advantage might be acquired through horizontal gene transfers. The main virulence mechanisms used by *Borrelia* and *Klebsiella* are discussed below and illustrated in figure 7.

2.5.1 *Borrelia*

Bacteria of the genus *Borrelia* belong to the Spirochaetaceae family and are Gram-negative bacteria characterized by a helical structure and didermic cell envelopes (Radolf et al., 2012). They can cause the vector-transmitted diseases Lyme borreliosis (LB) and relapsing fever (RF). LB, which is common in the northern hemisphere, is transmitted by *Ixodes* ticks and initially causes flu-like symptoms that can later develop into chronic symptoms affecting the joints, the nervous system and the heart (Schnarr et al., 2006). RF is commonly found in

temperate and tropical regions and causes recurrent fever episodes (Talagrand-Reboul et al., 2018). About 20 *Borrelia* species cause LB and these are generally referred to as the *B. burgdorferi* sensu lato complex (hereafter LB *Borrelia*). *B. burgdorferi* sensu stricto (hereafter *B. burgdorferi*) refers to a species within the sensu lato-complex.

Unlike many bacteria, the *Borrelia* genome is not known to encode any toxins or secretion systems (Fraser et al., 1997). This means that *Borrelia* relies on other virulence mechanisms to sustain infections, such as adhesion to host surfaces and immune evasion. By mainly trying to evade the host immune responses, rather than actively damaging the host cells, the symptoms of infection are thus caused only by the inflammatory responses produced by the host (Kerstholt et al., 2020). *Borrelia* enters the mammalian host through the tick bite, and tick saliva also contains proteins that help the bacteria evade the initial detection by the immune system, e.g. by inhibiting ROS production and by preventing complement, antibody-mediated killing and the release of antimicrobial peptides (Radolf et al., 2012).

Borrelia spreads in the host through the blood circulation and interacts with and adheres to host surfaces through adhesins located on the bacterial outer membrane. *Borrelia* encode many different adhesins and some of them have been shown to be essential for host infection while the function of others are still unknown (Brissette and Gaultney, 2014). The specific role of an adhesin during host infections is difficult to study since adhesins can have redundant functions or be expressed only in certain tissues, or at different stages of the infection (Petzke and Schwartz, 2015). Like other extracellular pathogens, *Borrelia* often attach to host cell surfaces or to the extracellular matrix, a network of proteins and carbohydrates that surrounds the cells. The adhesins on the bacteria can bind to different components e.g., collagen, fibronectin, glycosaminoglycans (GAGs), integrins and decorin (Brissette and Gaultney, 2014).

The outer surface of the *Borrrelia* contains an unusually large number of lipoproteins, but no lipopolysaccharides, like other Gram-negative bacteria (Petzke and Schwartz, 2015). The lipoproteins play vital roles in adhesion, and in immune evasion mechanisms like antigenic variation and complement evasion. The vls antigenic variation system is found in all LB-causing *Borrelia* and is an essential immune evasion mechanism needed for the long-term survival of LB *Borrelia* in mammalian hosts. It consists of the vlsE gene and a number of silent cassettes that contain variants of the central cassette found in vlsE (Norris, 2015). During mammalian infections, recombination events take place between these regions, ensuring that the antigenic surface of the protein is constantly changing and thereby prevents recognition by the host immune system. The complement system provides the first rapid defense against pathogens and can kill Gram-negative bacteria within minutes by forming a membrane attack complex that causes cell lysis (Heesterbeek et al., 2018). *Borrelia* use several

different mechanisms to evade detection by the complement, e.g., outer surface proteins called CRASPs (complement regulator-acquiring surface proteins), that bind to complement regulator proteins of the factor H family, which inhibits the complement pathway (De Taeye et al., 2013).

Borrelia are highly motile bacteria and can move very fast in the tissues, which enables them to avoid engulfment by phagocytes (Radolf et al., 2012). The LB *Borrelia* also have a rather small genome that is missing several vital metabolic pathways (Fraser et al., 1997). They are e.g. unable to synthesize amino acids, nucleotides and fatty acids, and rely solely on glycolysis for energy production (Kerstholt et al., 2020). The parasitic lifestyle of *Borrelia* enables them to obtain all necessary nutrients from their host. Although the *Borrelia* genome encodes fewer transporter proteins than other bacteria, many of them have a broad substrate-specificity and can import a wide range of substrates (Saier Jr. and Paulsen, 2000).

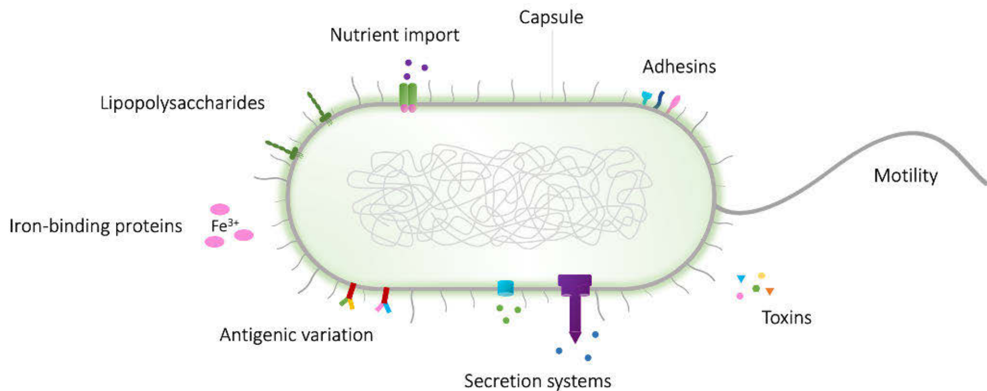


Figure 7. Main virulence mechanisms used by *Borrelia* and *Klebsiella* bacteria. Obtaining vital nutrients from the host can be achieved through specialized importers or by secreting proteins that capture specific nutrient, e.g., iron. For *Klebsiella*, a thick capsule plays a vital part in protecting the bacteria against recognition by the immune system. Antigenic variation is another way to escape detection by host immunity. Adhesins enable tissue colonization by adhering the bacteria to host cells or to the extracellular matrix. In *Borrelia*, flagella enable bacteria to move quickly through the host tissues and help them evade phagocytosis. Lipopolysaccharides and secreted toxins and effectors cause harmful effects in the host. See text for more details about which mechanisms are found in *Borrelia* and *Klebsiella* respectively.

2.5.2 *Klebsiella*

K. pneumoniae is considered an urgent threat to human health due to its increasing resistance to antibiotics (Paczosa and Mecsas, 2016). It is responsible for over 70 % of human *Klebsiella*-infections (Pitout et al. 2015) and causes infections of the respiratory and urinary tracts, as well as sepsis and abscesses of the liver (Broberg et al. 2014). Virulence genes and antimicrobial resistance genes are easily shared between *Klebsiella* bacteria, causing hypervirulent,

multi-resistant strains (Bengoechea and Sa Pessoa, 2019). The World Health Organization lists *K. pneumoniae* as one of the species that should be prioritized when developing new antibiotics (WHO, 2017).

K. pneumoniae is known as a stealth pathogen due to its skills in avoiding detection by the innate immune system (Bengoechea and Sa Pessoa, 2019). It can shield its PAMPs from being recognized by the immune system (Llobet et al., 2015), avoid phagocytosis (March et al., 2013) and prevent the effects of the complement system (Álvarez et al., 2000). Another characteristic feature of *K. pneumoniae* infections seems to be a reduction of the early inflammatory responses. If the body cannot properly initiate this early response, it results in a more severe infection. It has been shown that *K. pneumoniae* inhibits this response by blocking the activation of the NF-kB and MAPK pathways (Regueiro et al., 2011).

The best characterized virulence factors in *K. pneumoniae* are capsule, lipopolysaccharide, fimbriae and siderophores (Paczosa and Mecsas, 2016). The capsule is essential for *K. pneumoniae*, as strains lacking a capsule show a significant reduction in virulence (Cortés et al., 2002; Lawlor et al., 2006). The capsule protects the bacteria from the actions of the innate immune system, such as complement, phagocytosis and antimicrobial peptides (Paczosa and Mecsas, 2016). The thickness of the capsule seems to be more important than its composition when it comes to providing protection and this is evidenced by the thicker hypercapsule seen in hypervirulent *K. pneumoniae* strains (Patro and Rathinavelan, 2019). LPS are found in the outer membrane of Gram-negative bacteria and consist of an O antigen, a core oligosaccharide and lipid A. It is the major factor that protects *K. pneumoniae* against complement (Paczosa and Mecsas, 2016). Strains lacking a full-length O antigen are sensitive to complement killing, even if the bacteria is protected by a capsule, whereas strains containing an intact O antigen are resistant (Merino et al., 1992). Fimbriae mediate attachment to host cells and play an important part in biofilm formation, while siderophores are used to scavenge iron from the host environment (Paczosa and Mecsas, 2016).

3. Aims

The aim of this thesis was to make a detailed structure-function analysis of specific proteins involved in interbacterial competition and host-pathogen interactions in *Borrelia* and *Klebsiella* bacteria. The individual aims were:

Aim I – To elucidate the function, ligand-binding properties and structure of the *B. burgdorferi* Bmp proteins

The aim of this project was to determine the exact function of the *Borrelia burgdorferi* basic membrane proteins BmpA, BmpB, BmpC and BmpD and define their role in *Borrelia* infections. Although the Bmp proteins in the Lyme disease causing *B. burgdorferi* are known to play an important part during human infections, their function has remained unknown. The objective of publication I was to create 3D structural models of the Bmp proteins and to study structural differences between the proteins as well as their ligand-binding properties and evolutionary relationships. In publication II, the objective was to experimentally analyze the ligand-binding of BmpD and to determine its X-ray structure to give a structural explanation for its ligand-binding capacity.

Aim II – To reveal key residues involved in host cell adhesion in the *B. garinii* Dbp proteins

In this project, the aim was to determine the role of the *Borrelia garinii* DbpA and DbpB proteins in mediating attachment to host cells in the nervous system. *Borrelia garinii* is the main cause of Lyme disease in Europe and is often associated with neuroborreliosis. The adhesins DbpA and DbpB are vital for *B. garinii* virulence in mammalian hosts. The objective of publication III was to create 3D structural models of the Dbp proteins and to make a detailed analysis of how mutations in potential binding-site residues affect the function of the proteins.

Aim III – To predict the structures for *K. pneumoniae* VgrG4 and Sel1E and analyze their possible interaction sites

The aim of this project was to determine the structure and function of the *Klebsiella pneumoniae* VgrG4 protein and to study its role in host infection and intermicrobial competition. The highly antibiotic resistant *K. pneumoniae* causes a wide range of infections in the respiratory system, including life-threatening hospital-acquired infections. VgrG proteins are an integral part of the T6SS, which is used for secreting toxic effectors into host cells or competitors. In publication IV, the objective was to create 3D structural models for VgrG4 and its immunity protein, Sel1E, and to predict possible interaction sites between them.

4. Methods

Each section begins with a brief introduction to the methods used in this thesis and is followed by short descriptions of how the methods were used in each project. The main methods used are also summarized in figure 8.

4.1 Sequence analysis

The analysis of protein sequences is a fundamental part of bioinformatics and there are several widely used databases and tools for exploring protein sequences (Chen et al., 2017). From the amino acid sequence of a protein, it is possible to find information about its molecular mass, pI, stability etc., to find functional domains and motifs as well as to predict cell location, secondary structure and even possible interaction partners.

Amino acid sequences can be obtained from the UniProtKB (https://www.uniprot.org/) and from the NCBI protein database (https://www.ncbi.nlm.nih.gov/). The UniProtKB (Bateman et al., 2017) is a large repository of protein sequences containing detailed annotations about the sequences and links to other sequence analysis resources, e.g. InterPro (Finn et al., 2017), which classifies proteins into families and provides information on functionally important domains (Mitchell et al., 2019). The Basic Local Alignment Search Tool (BLAST) is an invaluable tool in bioinformatics (Altschul et al., 1990). The BLAST algorithm uses sequence databases to find local similarities between sequences and then calculates the statistical significance of the matches. A BLAST search with an unknown sequence can give information about the proteins function, evolutionary relationships, protein family and can also provide information about homologous proteins with a known 3D structure. The protein structures identified by BLAST searches can be retrieved from the Protein Data Bank (rcsb.org, Berman et al., 2000), which contains the 3D structures of biological macromolecules solved by X-ray crystallography, NMR or Cryo-EM.

All sequences used in this work have been retrieved from the UniProtKB or from the NCBI databases. The InterPro (Finn et al., 2017) sequence analysis tool was used to identify protein families and domains. Secondary structure predictions were made with PSIPred (Jones, 1999), and LipoP (Juncker and Willenbrock, 2003), and SignalP (Almagro Armenteros et al., 2019) was used to predict signal peptides. Localization predictions were made for the *K. pneumoniae* Sel1 proteins using CELLO (Yu et al., 2004), Psortdb (Peabody et al., 2016) and Phobius (Käll et al., 2007). The evolutionary conservation patterns of the Sel1 proteins were analyzed with ConSurf (Ashkenazy et al., 2016).

4.2 Sequence alignment

Sequence alignments are used to identify regions of similarity (conservation) between sequences and can give information about structural, functional or evolutionary relationships. Multiple sequence alignments align three or more sequences and is the starting point for a wide range of bioinformatics analyses (Chowdhury and Garai, 2017). Scoring methods are used to determine how similar the residues in a specific position in the alignment are. Two types of scoring are used for amino acid sequence alignments: an identity score, which indicates if identical residues are found in both positions and a similarity score, which also takes the physicochemical properties of amino acids into account. A substitution for a residue with similar properties will therefore score higher.

The sequence alignments have been made in the BODIL modelling environment (Lehtonen et al., 2004) using MALIGN, for multiple sequence alignments, and VERTAA for structure-based alignments. In publication I and III the multiple sequence alignments were aligned to the pre-aligned structure-based alignments using MALIGN. In publication IV the same method was used to make an alignment of the N-terminal part of VgrG4.

4.3 Modelling protein 3D structures

Protein function is closely linked to protein structure and changes in a single amino acid can completely alter or abolish the function of a protein if it is located in a structurally important region (Sadowski and Jones, 2009). Knowing the 3D structure of a protein is therefore often a necessary step towards elucidating its function. Protein 3D structures can be determined experimentally using X-ray crystallography, NMR and Cryo-EM, or predicted computationally using homology modeling, threading or *ab initio* modeling (Chandra et al., 2010). The experimental methods are considered highly reliable and produce high resolution structures, however, they are also time-consuming and expensive (Kc, 2017). The computational methods, on the other hand, are faster and inexpensive in comparison and function as an important complement to the experimental methods.

Related proteins, i.e., proteins that are homologous, often share a similar structure. In homology modeling this premise is used to model proteins based on an experimentally solved structure of a related protein (França, 2015). A sequence alignment between the structural template and the sequence of interest is the first, and also the most essential, step in the homology modeling process. A correct sequence alignment is a prerequisite for creating a good quality model (Haddad et al., 2020). The higher the sequence identity between the proteins the more likely that the model will be correct. However, as protein structures are more conserved than sequences (sequences may change without

resulting in structural changes) it is possible to create reliable models even though the sequence identity is low (>30%) (Illergård et al., 2009).

Modeller is one of the most commonly used homology modeling programs and it tries to find the most probable structure for a sequence by satisfying spatial restraints in the form of probability density functions (Šali and Blundell, 1993). C-alpha-C-alpha distances, main-chain N-O distances, main-chain and side-chain dihedral angles are restrained, and the program produces a model which fulfils these restraints as well as possible.

Protein threading is another frequently used template-based modeling method and it is used e.g. by I-TASSER (Yang et al., 2015), an online modeling server that has consistently ranked high in the CASP experiments where protein structure modeling methods are evaluated (Kryshtafovych et al., 2019). Threading methods align a protein sequence to proteins in the PDB and try to find matches that it can use as templates to produce a model. The final model is constructed by reassembling structural fragments based on the threading templates (Roy et al., 2011). Threading is normally used when there are no homologous proteins available in the PDB but the protein still shares part of the structure with proteins in the PDB (Xu et al., 2007). Threading does not require as high a sequence identity as homology modeling and can produce good alignments even with a sequence identity below 25 %.

If no structural template is available for a protein, *ab initio* modeling can also be used to predict its structure. This method predicts the native structure of a protein using only its amino acid sequence to find the conformation with the global energy minimum (Dorn et al., 2014). Until very recently this method has not been able to produce very accurate models. In the latest CASP experiments (https://predictioncenter.org/casp14/index.cgi, 22.1.2021) however, protein structures were for the first time predicted with an exceptionally high accuracy by the artificial intelligence method AlphaFold, developed by DeepMind (Jumper et al., 2020).

If structural templates are available, the first step in the homology modeling process is to create a structure-based alignment where the structural templates are superimposed and provide a structural framework on which to build a correct alignment for modeling. The sequence of interest is then aligned to the structure-based alignment and the structural features can be used to provide a more reliable alignment for modeling.

Homology models of the Bmp proteins, the Dbp proteins and the VgrG4 protein were made with Modeller (Šali and Blundell, 1993). 10 models were created for each protein and the model with the lowest energy according to Modeller objective function was chosen for further analyses. The modeling of VgrG4 was done in four separate parts. The N-terminal part (residues 1-563), the repeat

part (637-677) and the C-term part (678-739) where modeled with Modeller, while the helix-part (residues 564-636) was modeled using the modelling server I-TASSER (Yang et al., 2015). A structure-based alignment was made using the crystal structures of *Pseudomonas aeruginosa* VgrG1 (PDB ID: 4uhv (Spínola-Amilibia et al., 2016), 4mtk (Sycheva et al. to be published)) and the *Escherichia coli* v3393 protein (PDB ID:2p5z (Leiman et al., 2009)) and VgrG sequences of similar length to PaVgrG1 and to *K. pneumoniae* VgrG4 were aligned to the prealigned structure-based alignment, creating the alignment used for modeling. Due to the repetitive patterns found in the VgrG spike, we could use a repeated part of PaVgrG1 (residues 551-589) to model the repeat part of VgrG4. The C-terminal part was modeled based on *P. aeruginosa* VgrG1 residues 550-611.

Sel1D and Sel1E were modeled using the I-TASSER server and based on the homologs found in the BLAST search we specified that the program should use the *P. aeruginosa* Pa5087 protein (PDB ID: 5jkp) as the template.

4.4 Model analysis

The quality of protein models needs to be carefully evaluated before any further analyzes are done. A visual inspection of the model superimposed on the template is usually the first step. The RMSD (root mean square deviation) calculates an average distance between atoms in the two structures and gives an indication of the overall similarity (Kufareva and Abagyan, 2012). There are also several online protein quality evaluation servers that assess protein models based on e.g., stereochemical or electrostatic properties or similarities to known structures. All models were visually inspected and evaluated by superimposition with the template structures using the program VERTAA in the BODIL modeling environment (Lehtonen et al., 2004) and in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The models were also evaluated by the online model quality evaluation servers Qmean (Benkert et al., 2009), PROSAweb (Wiederstein and Sippl, 2007), ProQ (Wallner and Elofsson, 2003) and MODFOLD (Maghrabi and Mcguffin, 2017). The ABPS plugin in PyMOL was used to calculate the electrostatic potential of the Dbp (publication III) and Sel1 protein surfaces (publication IV).

4.5 Phylogenetic analysis

Phylogenetic analyses using DNA or protein sequences can be used to determine the evolutionary relationships between species or strains (Ajawatanawong, 2017). The evolutionary relationships are depicted in the form of a tree, which is calculated based on a multiple sequence alignment consisting of homologous sequences. The accuracy of the tree is entirely dependent on the quality of the sequence alignment. The methods used to calculate phylogenetic trees can be divided into two main groups: distance-based and character-based (Yang and

Rannala, 2012). In distance-based methods, a distance value is calculated based on the sequence similarities in the multiple sequence alignment and a distance matrix is then used to infer the phylogenetic tree. Neighbor joining is one of the most commonly used distance-based methods and is generally a very fast method but is not considered to be reliable for highly variable sequences. Among the character-based methods, maximum parsimony, maximum likelihood and Bayesian inference are the most frequently used methods (Ajawatanawong, 2017). These methods are considerably more time consuming than the distance-based methods because they calculate the phylogenetic tree directly based on sequence information. All possible trees that can be constructed based on the multiple sequence alignment are scored, and the best tree is determined based on the highest score. However, as it is not computationally practical to calculate all possible trees for larger sequence alignments, most methods use some form of heuristic approach to speed up the process (Yang and Rannala, 2012).

The phylogenetic tree (publication I) was generated using the Maximum Likelihood (ML) method in MEGA7 (Kumar et al., 2016). The Le and Gascuel (2008) substitution matrix (LG)(Le and Gascuel, 2008) with gamma distributions and invariant sites (G + I) was determined to be the most appropriate evolutionary model for the data set using the “Find Best DNA/Protein Models” function in MEGA and was used for the ML analysis. Complete deletion of gaps and missing data was carried out to exclude highly variable regions from analysis and bootstrapping (500 replications) was used to evaluate branch support. Bootstrapping is commonly used for evaluating the confidence of phylogenetic trees (Felsenstein, 1985; Hillis and Bull, 2010) and 500 replicates is the default setting in MEGA7 (Kumar et al., 2016).

4.6 Molecular dynamics simulations

Molecular dynamics (MD) simulations aim to imitate the motions of atoms in a molecular structure set in an environment that mimics the normal setting of that molecule (Patodia, 2014). For proteins, this means that the motions of a protein structure are simulated in a water-filled environment that imitates its normal cellular environment. The system is held at a defined temperature and pressure and ions can be added to neutralize the charge on the protein. A force field is used to calculate the potential energy of the system. MD-simulations can be used for a wide range of purposes, e.g. to study protein stability or folding, to analyze the effects of mutations or to observe conformational changes or ligand-binding (Hollingsworth and Dror, 2018).

The MD simulations of the Dbp models (publication III) were done with Desmond Molecular Dynamics System (2019-4) in Maestro (Bowers et al 2006). The proteins were solvated in an octahedral box, with 20 Å distance between the solute surface atoms and the edge of the box, using a TIP3P water model (Jorgensen 1983). Cl^- atoms were added to neutralize the systems. A constant

temperature of 300 K (maintained by the Nosé-Hoover chain thermostat) and a constant pressure of 1.01325 bar (maintained by the Martyna-Tobias-Klein barostat) was upheld using the OPLS3e force field (Harder 2016). The default equilibration protocol in Desmond was used to relax the systems and the simulations were run for 200 ns. All simulations were done in triplicates by randomizing the initial velocities and the resulting trajectories were analyzed with the Simulation Interactions Diagram program in Desmond.

4.7 Visualization and structural analysis

Manual visualization and analysis of proteins structures and their potential interactions are one of the most important ways of evaluating models, simulations or dockings. By superimposing models on their structural templates, it is possible to easily observe similarities and identify differences. Individual frames of an MD-simulation and the different poses of docked molecules can be analyzed in great detail. Visualization software allows the user to show individual amino acid residues in different formats, to visualize potential bonds and to measure atomic distances.

Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) was used for visual inspection of the models and for creating high resolution pictures. Labels were added in Microsoft® PowerPoint or in the GNU Image Manipulation Program. ESPript 3.0 (Robert and Gouet, 2014) was used to prepare the alignment figures. Graphical figures were created in PowerPoint.

4.8 Experimental work

Except for the BmpD crystallization described below, all experimental work in this thesis was performed by our collaborators in their laboratories. The experiments are explained in detail in the Material and Methods sections of the publications.

4.8.1 BmpD crystallization

The sitting-drop vapor diffusion method was used to obtain crystals of rBmpD. 9.6 mg/ml protein in 50 mM Tris/HCl pH 8.0 buffer was mixed 2:1 with the reservoir solution of 0.2 M sodium chloride, 0.1 M Tris, 20% (wt/vol) polyethylene glycol 6000 (pH 8.0), supplemented with 15% 2-methyl-2,4-pentanediol (MPD) as cryoprotectant. The crystals were observed after 5 days.

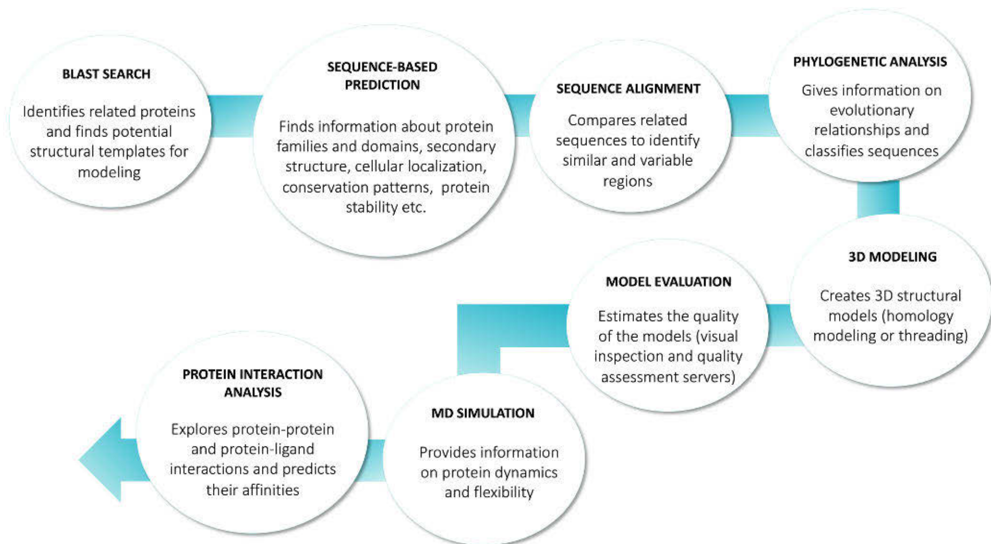


Figure 8. Summary of the structural bioinformatics methods used in this thesis.

5. Results and Discussion

5.1 *Borrelia burgdorferi* – BmpA, BmpB, BmpC and BmpD

5.1.1 Introduction

The *B. burgdorferi* Bmp proteins were first characterized in the beginning of the 1990s when antibodies against BmpA (then called P39) were detected in serum samples from human LB patients (Simpson et al., 1990). The gene was located in the single chromosome found in *B. burgdorferi* and soon three additional genes (*bmpB, bmpC* and *bmpD*) were discovered in the same location, all homologous to *bmpA* (Aron et al., 1994; Ramamoorthy et al., 1996; Simpson et al., 1994). Antibodies against BmpA are specific and can be used for diagnosing LB patients (Verma et al., 2009). Antibodies against BmpB and BmpD have also been found in LB patients, whereas BmpC is shown to be continuously expressed during infection in mice (Ting Liang et al., 2002). Thus, the Bmp proteins are expressed during infections in humans, but their exact function is unknown. The aim of this work was to determine the structure and function of the Bmp proteins and to define their role in LB infections.

5.1.2 The *B. burgdorferi* Bmp proteins are substrate-binding proteins

The four Bmp proteins of *B. burgdorferi* B31 have a length of 339 (BmpA), 341 (BmpB, BmpD) and 353 (BmpC) amino acids and share sequence identities in the range of 37-50 %. Their expression during *Borrelia* infections indicate that the Bmp proteins play an essential part in the infection process. However, as there has been conflicting reports about their location and function, their exact function has remained unknown. Verma et al. (2009) predicted that the Bmp proteins are adhesion proteins binding to laminin. However, since BmpB, BmpC and BmpD are predicted to be located on the inner membrane of the periplasmic space, this would preclude a role in adhesion (Dowdell et al., 2017). Others have predicted that the Bmp proteins are ABC-transporter proteins (Gherardini et al., 2010; Kanehisa and Goto, 2000). Our own sequence analyses with InterPro also predicted that the Bmp proteins contain an" ABC transporter substrate-binding protein PnrA-like" (IPR003760) domain and a "Periplasmic binding protein-like I" (IPR028082) domain. The Bmp proteins were also predicted by LipoP (Juncker and Willenbrock, 2003) to have a lipoprotein signal sequence at the N-terminal part. Lipoprotein signal sequences are used by bacteria to ensure that proteins are positioned in their correct location in a membrane (Zückert, 2014). The signal sequence consists of a region of approximately 20 amino acids, from the first amino acid (Met) to the first cysteine residue (Zückert, 2014). The proteins are attached to the membrane via a diacyl glyceryl group, which is bound to the sulfhydryl group of the cysteine, and an acyl group attached to the amino group

(Buddelmeijer, 2015). The signal sequence is then cleaved from the protein before it is attached to the membrane.

5.1.2.1 Analyses of the Bmp structure and models

A BLAST search (Altschul et al., 1990) was done to find homologous proteins with a known 3D structure. The related protein PnrA in *Treponema pallidum* (hereafter Tp) has a sequence identity of 27.5-31.3 % to the Bmp proteins and has been shown to be an ABC-transporter substrate-binding protein that binds to purine nucleosides (Deka et al., 2006). This protein was subsequently used as a structural template for modeling the Bmp proteins.

Homology modeling, the methods used for modeling the Bmp proteins, is a frequently used protein structure prediction method, which is based on the assumption that proteins with similar sequences also have similar structures (França, 2015; Rost and Sander, 1996). The function of a protein is dependent on its structure and protein structures are thus more conserved than sequences in order to preserve protein functions. It is therefore possible to model a protein based on a related 3D structure even though the sequence identity is low (Illergård et al., 2009). Quality assessments, performed by visual inspection and by assessment servers, showed that the models of the Bmp proteins were of high quality. The root mean square deviations (RMSD) between the Bmp proteins and PnrA were in the range of 0.42-0.5 Å, indicating highly similar structures. The quality assessment servers also gave high scores to the models, further supporting their quality.

In parallel with the modeling, the crystal structure of BmpD was solved (Publication II, Table 1). The crystal structure superimposes on the Tp PnrA structure with an RMSD of 1 Å, indicating a very high structural similarity despite the low sequence identity of 27.8 %. In the crystal structure, BmpD was also, surprisingly, bound to a purine nucleoside, which had not been added during the crystallization setup. As Tp PnrA is also known to bind to purine nucleoside, this further supports the notion that PnrA and the Bmp proteins have a similar function as well as a similar structure.

The Bmp proteins share the typical fold of substrate-binding proteins (SBPs), which consist of two alpha/beta domains linked by three connecting loops (figure 9) (Scheepers et al., 2016). A substrate-binding site is formed in the cleft between the two domains and residues from both domains bind to the substrate. SBPs can have many different functions and bind to several different substrates, but they all have a highly conserved fold, despite having low sequence similarities. According to the classification system, which divides the SBPs into seven structural clusters (Scheepers et al., 2016), the Bmp proteins belong to the B-I cluster. This group consists mainly of proteins binding to sugars, alcohols and autoinducer 2. When a substrate binds to the SBP it induces a conformational

change to a closed formation, which is recognized by the TMDs and triggers hydrolysis of ATP and opening of the translocation channel (see figure 4).

The BmpD model was shown to be highly similar to the BmpD crystal structure and differences were found mainly on the surface of the proteins. The structures superimpose with an RMSD of 1 Å. Helices I and J are shorter in the model than in the crystal structure and there are some small differences in the position of surface loops. Especially the region surrounding the ligand binding site is very similar in both of them, except for some differences in water molecules surrounding the ligand, reflecting the slightly different binding modes for inosine and adenosine (see more details below).

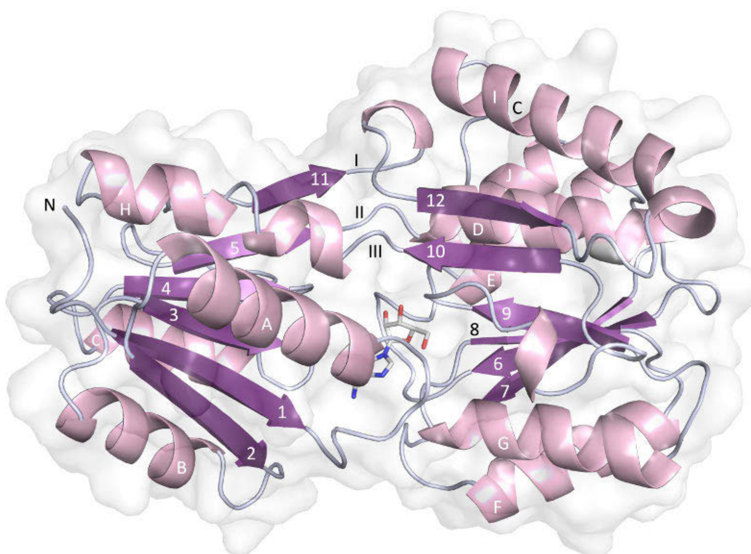


Figure 9. Crystal structure of *B. burgdorferi* BmpD bound to adenosine. The central beta strands in the center of each domain are shown in purple and numbered 1-12. The surrounding alpha helices are in light pink and marked A-I. The ligand, adenosine, is shown as white sticks in the binding site formed in the cleft between the domains. The three loops (I-III) shown above the ligand connect the two domains. Figure adapted from Publication II.

5.1.3 Structural characterization of the Bmp ligand-binding site

Nitrogenous compounds like nucleobases and nucleosides are needed by all organisms to form the nucleotides needed as building blocks for DNA, RNA and ATP. Nucleosides consist of a ribose part connected to a base part. In purine nucleosides, the base part is made up of a fused imidazole and pyrimidine ring (see figure 3), compared to the single pyrimidine ring found in pyrimidine nucleosides. Using LC-MS analysis, the purine nucleoside in BmpD was shown to be adenosine. However, ligand-binding assays demonstrated that BmpD also

binds to inosine, and therefore, it seems likely that the Bmp proteins are capable of binding to several similar nucleosides. In fact, Tp PnrA was also crystallized with several different purine nucleosides (Deka et al., 2006) and many other SBPs are known to bind to several structurally similar ligands (Davidson et al., 2008b). Our structural analyses indicate that conformational changes and additional water molecules can compensate for differences in the nucleoside structures and ensure that corresponding interactions are formed between BmpD and the ligand (see more details in section 5.1.3.2).

The inosine molecule bound to Tp PnrA was included in modeling the Bmp proteins, and this enabled us to make a detailed study of the substrate-binding properties of the ligand-binding sites of the Bmp models. The crystal structure of the *A. pernix* K1 solute-binding protein (PDB ID: 4pev, to be published), bound to adenosine, was also included in the analysis.

5.1.3.1 Inosine binding-site comparison

The ribose part of inosine in Tp PnrA forms hydrogen bonds with the side chains of D108, D238, K260 and with the main chain nitrogen of G212 (figure 10A), and these interactions are conserved also in the *A. pernix* K1 solute-binding protein and in the Bmp proteins, with a few exceptions. D238 is replaced by N241 in BmpC (figure 10E) but can still form the hydrogen bonds to O3' and O5' like the aspartate in PnrA. D108 is E98 in BmpD (figure 10F), which can form the same interactions as aspartate despite the longer side chain.

The residues binding to the base part of the purine nucleoside are conserved in BmpA, BmpB and BmpD, but not in BmpC. F36 and F186 in PnrA form aromatic stacking interactions with the imidazole and pyrimidine rings on the purine base. D27 forms hydrogen bonds with N1 and O6 and is considered a crucial residue for purine binding in Tp PnrA (Deka et al., 2006). The substrate-binding site in the BmpC model differs significantly from the other Bmps with regards to the residues interacting with the purine base part. F36 and F186 in PnrA are replaced by Y29 and L191 in BmpC (figure 10E). While the stacking interaction with Y29 is maintained, L191 is not capable of forming aromatic stacking interactions. BmpC also has a positively charged histidine (H21) instead of the negatively charged aspartate at D27 in PnrA, and this residue is furthermore turned away from the binding site, making hydrogen bonding to N1 and O6 unlikely.

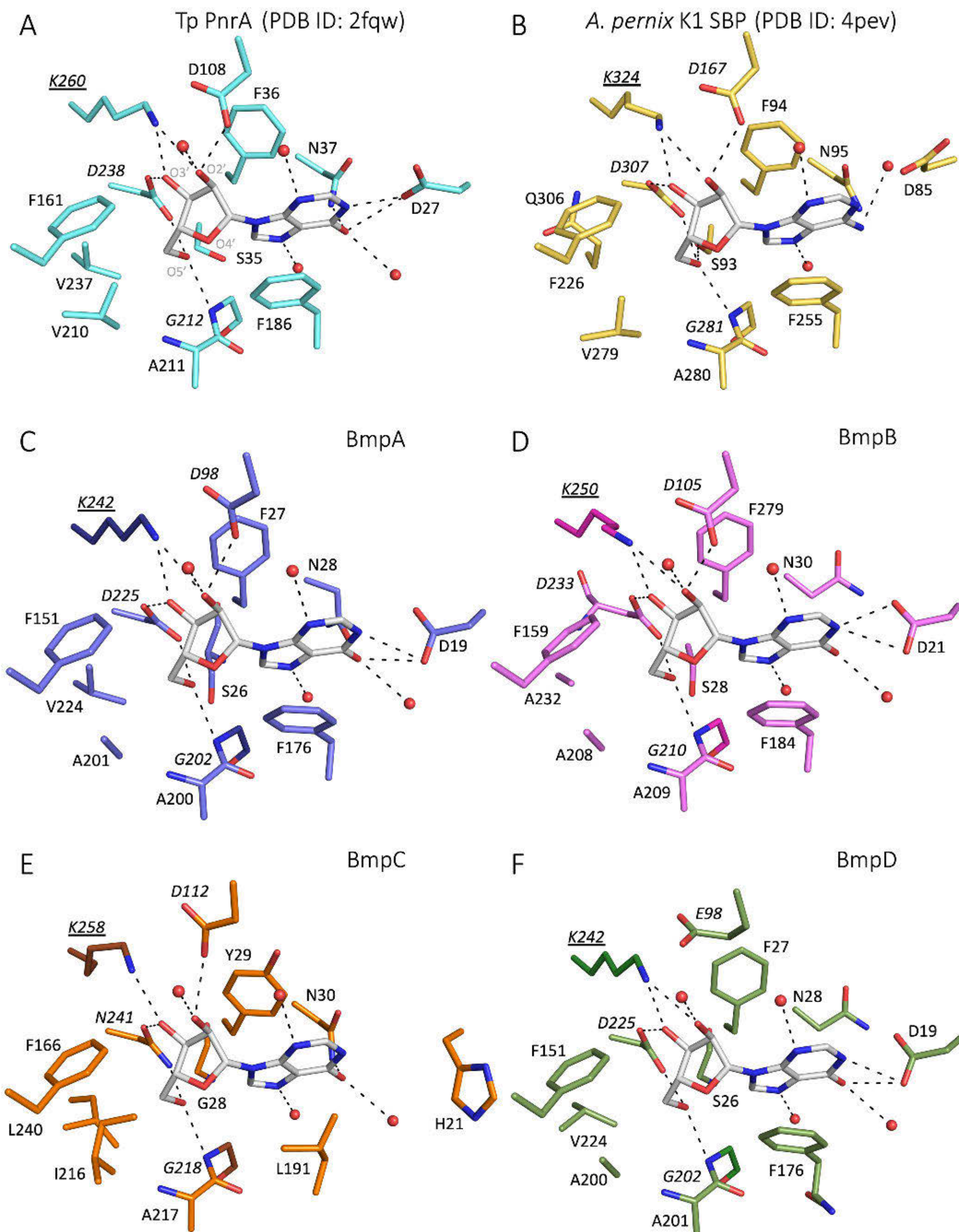


Figure 10. The ligand-binding site in the crystal structures of the *T. pallidum* PnrA-inosine complex (PDB ID: 2fqw, (Deka et al., 2006))(A), the *A. pernix* K1 solute-binding protein (SBP) in complex with adenosine (PDB ID: 4pev, to be published) (B) and the Bmp models in complex with the potential ligand inosine, BmpA (C), BmpB (D), BmpC (E) and BmpD (F). The atom names for the oxygens in the ribose part are marked in grey in A (O2'-O5'). The name of the completely conserved lysine is underlined and residues that bind to the ribose part are in italics. Figure adapted from Publication I.

BmpA and BmpB share 50.7 % sequence identity and the residues in their ligand-binding site are completely conserved with those of PnrA. BmpD is highly similar to BmpA and BmpB (47.2 and 44.0% sequence identities, respectively) whereas BmpC share only 35.5-37.8 % sequence identity with the other Bmps, and the residues that bind to the base part of the nucleoside differ notably from the other Bmps and PnrA. Thus, the residues binding to the ribose part of the nucleoside are highly conserved in all the proteins or conservatively substituted (BmpC, BmpD) so that the interactions are maintained despite the change in residue. In BmpC the residues binding to the purine base part are considerably less conserved, which suggests that BmpC could prefer other nucleosides than PnrA and the other Bmps and therefore might have a different function.

5.1.3.2 Comparison of adenosine and inosine binding

The purine nucleosides adenosine and inosine are identical except for position 6 on the base part, which has an amino group in adenosine and a carbonyl group in inosine (see figure 3). A structural model for inosine-bound BmpD was created using the PnrA-inosine complex structure (PDB ID: 2fqw (Deka et al., 2006)) in order to make a detailed comparison of the structural differences of adenosine and inosine binding in BmpD and PnrA.

The adenosine-binding sites in BmpD and PnrA differ mainly in the two loops that flank the binding site (figure 11A). In loop 1, S86 and F87 are replaced by F76 and R77 in BmpD (Figure 11B-C). D27 in loop 2 of PnrA forms hydrogen bonds with the backbone nitrogen of F87 (figure 11C), while D19 in BmpD is turned away from the binding site and forms ionic interactions with R77 (figure 11B). Loop 2 in PnrA is one residue longer than in BmpD and thus extends more deeply into the binding site (figure 11A). BmpD on the other hand has a more extensive water-mediated hydrogen-bonding network that connects the base part of the nucleoside with residues in loop 2.

Inosine binding to PnrA results in a conformational change in loop 2, which forces S28 to turn away from the binding site and its position is taken by water molecule 6 (figure 11E). Furthermore, water molecule 4 is excluded, which allows a direct interaction between the carbonyl oxygen of inosine and D27 (figure 11E). Loop 2 remains unchanged in the BmpD-inosine model as S28 is replaced by G20 in BmpD and the carbonyl oxygen of inosine forms hydrogen bonds with water molecule 6 and the main-chain oxygen of D19 (Figure 11D). Thus, the slight changes in residues and their positions are compensated for by alternative, but equally functional, interactions and is an excellent example of how protein structure and function can be maintained despite changes in vital residues.

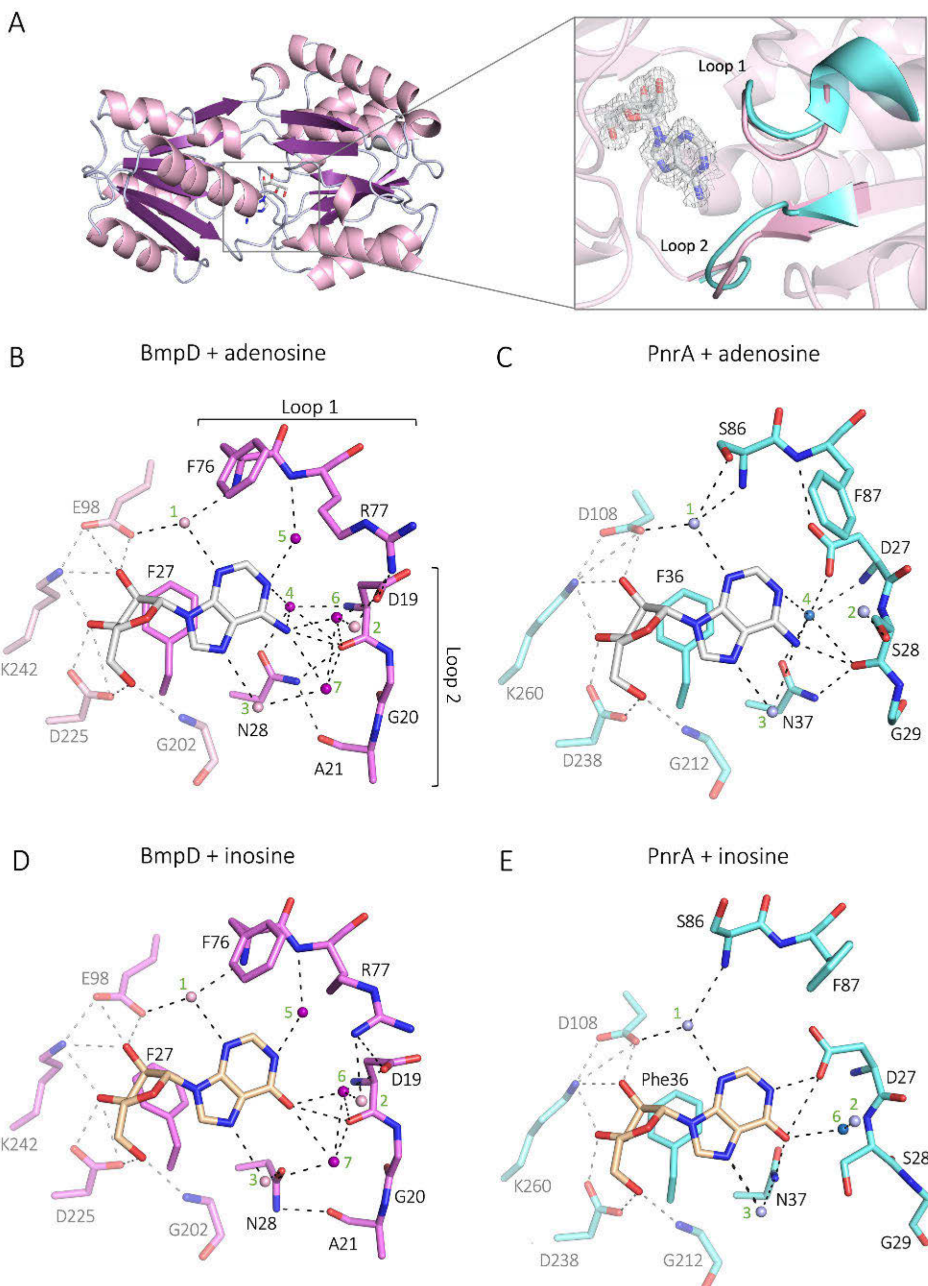


Figure 11. A. Crystal structure of BmpD bound to adenosine. The close-up shows the electron density map for the ligand and the position of the two loops whose residues bind to the base part of the purine nucleoside. B-E. Comparison of the ligand-binding site of BmpD (B), PnrA bound to adenosine (PDB ID: 2fqy) (C), the model of the BmpD-inosine complex (D), and PnrA bound to inosine (PDB ID: 2fqw) (E). All the potential hydrogen bonds are shown as dashed lines. The interactions of the base part are highlighted to display the differences, whereas the ribose part forms identical interactions in all the structures. Adenosine is shown as white sticks and inosine as wheat sticks, water molecules are shown as spheres (1-7); lighter colored spheres are completely conserved, and darker spheres differ between the structures. F176 (BmpD) and F186 have been omitted for clarity. Figure adapted from Publication II.

5.1.4 The Bmp proteins are conserved in *Borrelia* species

The majority of protein sequences are given a putative functional annotation based purely on sequence similarity searches, where the function of the top hit is automatically transferred to the new sequence (Brown and Sjölander, 2006). This frequently leads to annotation errors and to generalized and unspecified annotations. This was the case for the Bmp proteins, as a search of the UniProtKB database (Bateman et al., 2017) revealed that many Bmp sequences were annotated as "basic membrane protein", "ABC transporter substrate-binding protein" or "nucleoside-binding protein". Another group of proteins were annotated as "exported protein". By inferring a phylogenetic tree of the *Borrelia* Bmp sequences we aimed to improve the classification of the Bmp proteins and to identify functional differences between them. Furthermore, we wanted to determine if there would be a clear distinction between the Bmp proteins of LB and RF *Borrelia*.

The Maximum Likelihood (ML) method is frequently used to study sequence evolution (Yang and Rannala, 2012) and was therefore chosen for our analysis. ML gives a phylogenetic tree with the highest likelihood of producing the original multiple sequence alignment. The multiple sequence alignment in our case consisted of Bmp and exported protein (ExP) sequences from known LB and RF *Borrelia* strains, selected from the UniProtKB (Publication I, figure S1). The ExP sequences were included in the analysis in order to clarify their relationship to the Bmp proteins and, if possible, determine their function.

The ML-tree forms four main branches which further diverge into separate branches for the different Bmp proteins and, with some exceptions, the LB and RF *Borrelia* proteins then form distinct sub branches (figure 12). The first branch contains the ExP/BmpC proteins, the second branch the BmpD/BmpA2 proteins, the third branch the BmpB proteins and the last branch contains the BmpA proteins. The main branches are supported by high bootstrap values, except for the branch that divides into the ExP/BmpC and the BmpD/BmpA2 branches (bootstrap value 44). Low bootstrap values usually suggest that sequences have diverged during a relatively short evolutionary time frame (Soltis and Soltis, 2014, 2003).

The first main branch is split into separate branches for the ExP and the BmpC proteins and these are further split into sub branches for LB and RF *Borrelia* sequences. The ExP proteins are found within the same branch as the *T. pallidum* RfuA protein (Deka et al., 2013), which is also an ABC type substrate-binding protein and shares their typical structure, but which binds to riboflavin instead of purine nucleosides. It is thus involved in a completely different metabolic pathway than the nucleoside-binding proteins. The binding site in RfuA is located in the cleft between the two domains, as in other substrate-binding proteins, but requires a very different binding site to accommodate the riboflavin molecule,

which consist of a three-ringed isoalloxazine attached to a ribityl chain. Thus, since the ExP proteins are more closely related to RfuA than to the Bmp proteins they could likewise function as riboflavin-binding proteins.

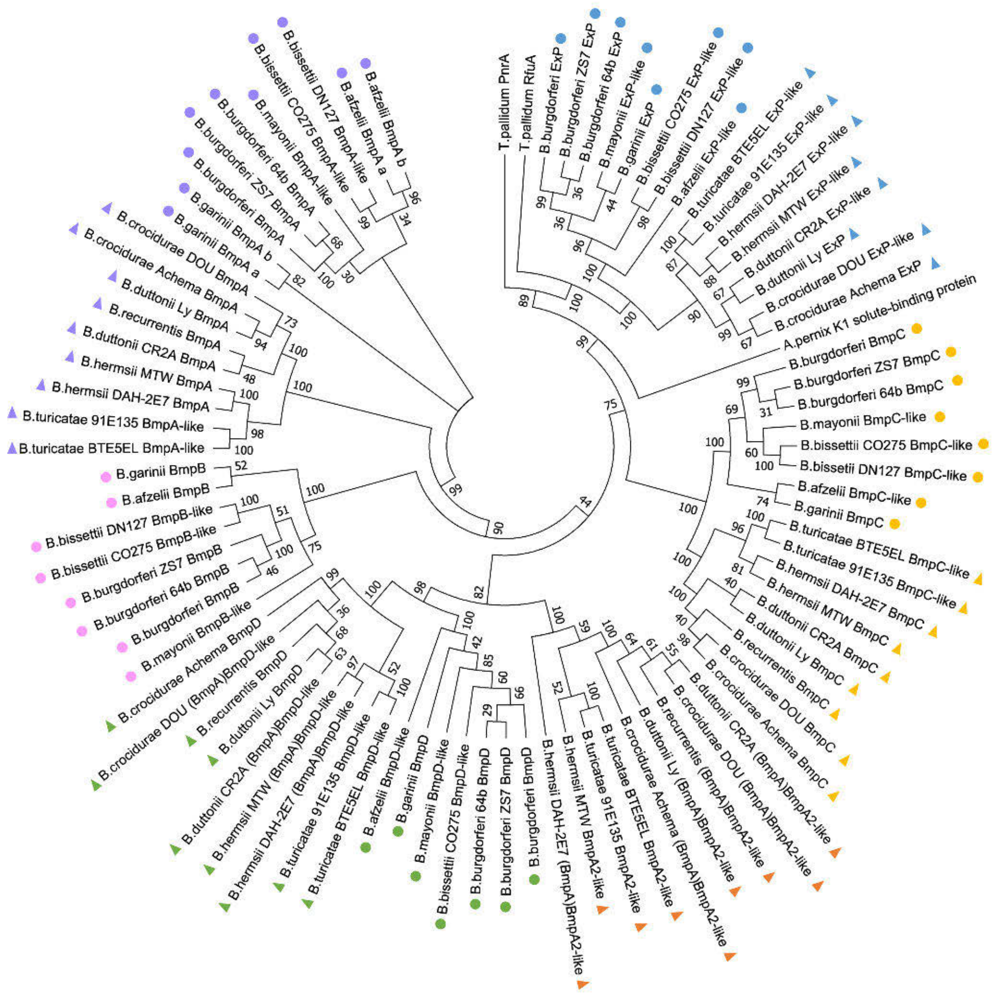


Figure 12. Maximum likelihood tree of the *Borrelia* Bmp and ExP proteins. Each main branch is individually colored and LB Borrelia species are marked with circles and RF Borrelia with triangles. Bootstrap values show branch support. Strain names are included if there is more than one strain per species. Proteins have been named BmpX-like if their names are not specified in the UniProtKB database, and if the UniProtKB names differ from how the tree classifies the protein, the UniProt name is found in parenthesis. Figure from Publication I.

The BmpC proteins are located in the same main branch as the ExP proteins and share a more recent common ancestor with them than with the rest of the Bmp proteins. However, the binding site of BmpC, though different from the other Bmp proteins, is even more different than the RfuA binding site, which indicates that the BmpC proteins are not riboflavin-binding proteins. Sequence analysis of

the BmpC proteins, as well as the structural analysis, show that the residues binding to the ribose part of the nucleoside are conserved whereas the residues surrounding the purine base part show considerable variation. In the position corresponding to H21 (figure 10E), BmpC proteins from other LB *Borrelia* have either histidine or asparagine, whereas in the RF *Borrelia* this position has proline or serine. In all other Bmps this position has a highly conserved aspartate that forms interactions with the base part of the ligand. Histidine, asparagine or serine could possibly form similar interactions with the ligand, but proline could not. L191 is a leucine or an arginine in LB *Borrelia* and glutamate or aspartate in RF *Borrelia*. These residues cannot form the stacking interaction with the base part of the ligand, as phenylalanine does in the other Bmp proteins. And the charged glutamate or aspartate in this position of RF *Borrelia* suggests that they could form ionic interactions or hydrogen bonds to the ligand molecule. N30 is conserved in all LB *Borrelia* but in RF *Borrelia* the corresponding residue is phenylalanine or leucine. Taken together, it is not clear how these variations in the ligand-binding site affect the function of the BmpC proteins, but it seems likely that the substrate they bind to could contain a ribose ring, or a ribose-like structure, as the residues surrounding that part are highly conserved. The variations between the LB and RF *Borrelia* also indicate a possible functional difference between these two groups, which could reflect the metabolic distinctions between the LB and RF *Borrelia.* RF *Borrelia* has a complete purine salvage pathway while the pathway in LB *Borrelia* is missing some key enzymes (Pettersson et al., 2007).

The RF *Borrelia* contains an additional Bmp protein, which is not found in the LB *Borrelia*. This protein has been named BmpA2 by Lescot et al. (2008) but in our phylogenetic tree the BmpA2 proteins are located in the same branch as the BmpD proteins. However, the ligand-binding residues of the BmpA2 proteins differ from the BmpD proteins by having a glycine instead of N28 (Publication I, figure S2). The asparagine residue in this position is conserved within all other Bmp proteins included in this study (with the exception of the BmpC proteins of RF *Borrelia*, which have a phenylalanine or a leucine instead). The asparagine side chain interacts with the purine nucleoside in the hydrogen-bonding network formed between the base part of the ligand and loop2 of the protein (figure 11), and as glycine lacks a side chain it cannot form the same interactions. Thus, these results indicate that BmpA2, like BmpC, could interact with a slightly different ligand than the other Bmp proteins.

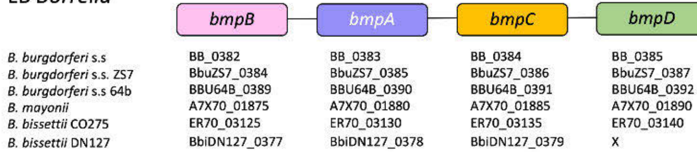
No RF *Borrelia* sequences are found in the BmpB branch, which is in agreement with the analyses done by Lescot et al. 2008, where no BmpB proteins were identified in the RF *Borrelia* (Lescot et al., 2008). For the BmpA proteins the sub branches are completely separate and do not form a common branch before diverging into individual branches for LB and RF *Borrelia*.

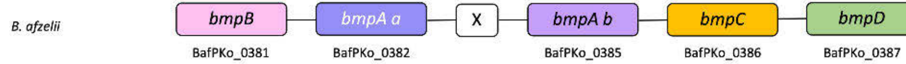
5.1.4.1 Genomic comparison and gene expression

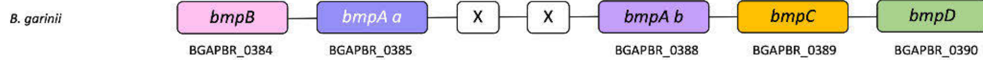
The *B. burgdorferi* Bmp proteins are expressed simultaneously *in vitro* but to varying degrees (Dobrikova et al., 2001). BmpA is expressed at a considerably higher level than the other Bmp proteins, and especially BmpC is expressed in very low amounts. BmpB is transcribed together with BmpA whereas the other proteins have their own transcriptional start sites (Ramamoorthy et al., 2005). This suggests that the BmpA protein might have a more important role in the purine salvage pathway or that it is involved in transporting a substrate that is in high demand by the bacterium.

The differences between the Bmp proteins of LB and RF *Borrelia* are also noticeable at the genome level. In both groups, there are four different *bmp* genes but their order in the genome differs (figure 13). In *B. afzelii* and *B. garinii* additional *bmp* genes are found, which are most likely the result of gene duplication events. In both *B. afzelii* and *B. garinii* there is an additional BmpA gene, which clusters together with the other BmpA proteins in the phylogenetic tree and has a high sequence identity (> 94 %) to the canonical BmpA protein in each species. To distinguish them from the BmpA2 proteins, which forms a group of its own, we labeled them BmpA a and BmpA b. The BmpA b protein of *B. afzelii* contains a serine instead of G202 and the *B. garinii* BmpA b protein has an asparagine instead of D19, however, neither change will likely impact the binding site properties. Since G202 forms a hydrogen bond to the ribose part through its backbone nitrogen, a change to serine would not affect this. However, the larger side chain of serine could impact the binding pocket if the side chain protrudes into it. The side chain nitrogen and oxygen of the asparagine would still be able to form the same hydrogen bonds as the aspartate.

LB *Borrelia*







RF *Borrelia*

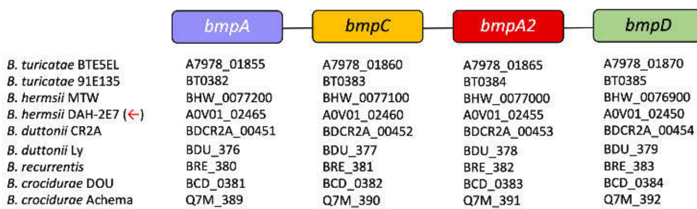


Figure 13. Gene order of the *Borrelia* bmp genes, based on the phylogenetic analysis. X indicates fragments or missing sequences in the NCBI database (Agarwala et al., 2017). The red arrow indicates that the order of the genes is reversed compared to the others. Figure adapted from Publication I.

5.1.5 The Bmp proteins play a role in the purine salvage pathway

Having determined the structures and ligand-binding properties of the Bmp proteins, we also studied their role in a wider cellular context. *B. burgdorferi* is an auxotrophic bacterium and cannot synthesize amino acids, fatty acid, vitamin cofactors or nucleotides *de novo* (Fraser et al., 1997; Gherardini et al., 2010). Bacteria that lack the capacity to synthesize important nutrients *de novo* have to rely on salvage pathways to obtain the necessary substances from the host environment.

Purine salvage pathways are used to re-form purine nucleotides from nucleobases and nucleosides obtained from the host (Warner et al., 2014). The purine salvage pathway in *B. burgdorferi,* however, is missing some of the essential enzymes needed (Pettersson et al., 2007). In particular, the vital enzyme ribonucleotide reductase (RNR) is missing in *B. burgdorferi*. This enzyme, which catalyzes the reduction of ribonucleotides into their corresponding deoxyribonucleotides, is considered essential for DNA synthesis in all organisms (Gherardini et al., 2010). *B. burgdorferi* compensates for the lack of this enzyme by relying on access to host deoxyribonucleotides but can also

make use of 2′-deoxyribosyltransferase, an enzyme that can transfer deoxyribose moieties from host-derived nucleosides to a nucleobase (Lawrence et al., 2009).

Borrelia obtain the required nucleobases and nucleosides from the host environment and convert these into nucleotides. Our results show that BmpD play a part in the uptake of nucleosides in *B. burgdorferi*, by transporting purine nucleosides in the periplasmic space to ABC-transporters at the inner membrane (figure 14). Purine nucleoside monophosphates from the host are first converted to nucleosides by a nucleotidase enzyme on the outside of the bacterial cell (Lawrence et al., 2009) before they pass through into the periplasmic space via outer membrane porins, such as p66, which allows the diffusion of small hydrophilic molecules (Barcena-Uribarri et al., 2010). BmpD is anchored to the inner membrane by a fatty acid chain (Dowdell et al., 2017) and binds to free purine nucleosides and transports them to ABC transporters (BB0677 to BB0679). The nucleoside is imported into the bacterial cytoplasm by the ABC transporter, and there deoxynucleotide kinase (BB0239) converts it back to a nucleoside monophosphate (Gherardini et al., 2010). Adenylate kinase (BB0417) and nucleoside diphosphate kinase (BB0463) then add additional phosphates and forms nucleoside diphosphates, and finally nucleoside triphosphates, which can subsequently be used as building blocks for RNA (Jewett et al., 2009).

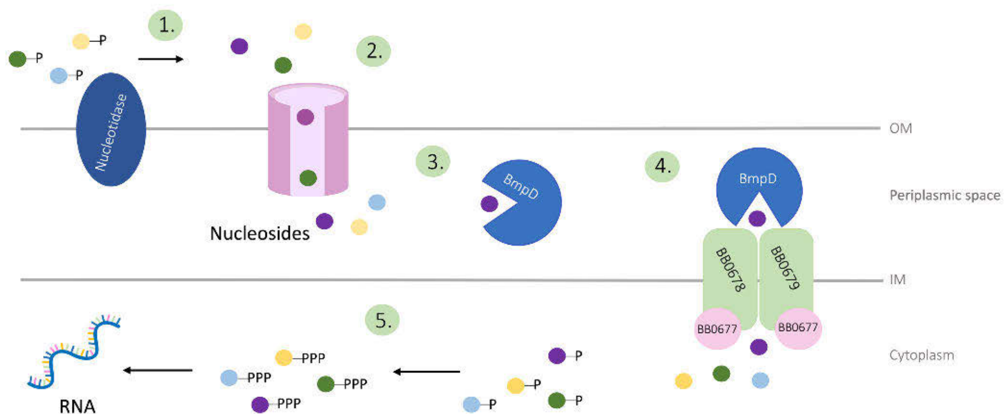


Figure 14. Schematic view of the purine salvage pathway in *B. burgdorferi*. Step 1, nucleotides ((i.e., GMP, IMP, and AMP) are converted into nucleosides (guanosine, inosine, and adenosine) by a nucleotidase enzyme by removing the phosphate group. Steps 2-3, The nucleosides are transported into the periplasmic space through the outer membrane and bind to SBPs, like BmpD. Step 4, the SBP protein transports the nucleoside to a membrane-bound ABC transporter system (BB0677 to BB0679), where they are transported through the inner membrane into the cytoplasm. Step 5, deoxynucleotide kinase (BB0239) adds a phosphate group to the nucleoside, reforming a nucleoside monophosphate. Adenylate kinase (BB0417) and nucleoside diphosphate kinase (BB0463) then add additional phosphates and first form nucleoside diphosphates and finally nucleoside triphosphates (ATP and GTP), which can be incorporated into RNA. Figure adapted from Publication II.

5.2 *Borrelia garinii* – DbpA and DbpB

5.2.1 Introduction

B. garinii (Bg), belonging to the *B. burgdorferi* sensu lato complex, is one of the main species that cause LB in Europe and is often associated with neuroborreliosis (Stanek and Strle, 2018). The *B. garinii* decorin-binding proteins A and B (DbpA, DbpB) are adhesins expressed during mammalian infections and mediate attachment to the proteoglycans decorin and biglycan (Brissette and Gaultney, 2014; Lin et al., 2017). Different *Borrelia* genospecies bind to the proteoglycan GAG chains with different affinities and this affects the tissue tropism of the spirochetes (Benoit et al., 2011; Lin et al., 2014; Salo et al., 2011). Dbp proteins form electrostatic interactions with the negatively charged GAG chains on the proteoglycans (Gandhi and Mancera, 2008). Brown et al. (1999) have shown that three lysine residues are critical for GAG-binding in *B. burgdorferi* and form the so called canonical GAG-binding site (Brown et al., 1999).

The Dbp proteins were first discovered in *B. burgdorferi* as surface proteins involved in bacterial adherence to decorin (Guo et al., 1995). They were characterized as surface-exposed proteins of the MSCRAMM (microbial surface component-recognizing adhesive matrix molecule) family (Guo et al., 1998) and shown to be expressed *in vivo* during infections (Hanson et al., 1998). Antibodies against DbpA can inhibit the growth of *B. burgdorferi* both *in vivo* and *in vitro* and immunization with DbpA gives complete protection. DbpA and DbpB were shown to mediate attachment to the extracellular matrix (Guo et al., 1998) and to be able to attach to mammalian cells (Fischer et al., 2003). The Dbp proteins seem to be important for the early stages of bacterial dissemination, which is partly mediated through the lymphatic system (Imai et al., 2013), and have also been shown to be required for arthritis development (Salo et al., 2015).

Since Bg is known to cause neurological symptoms in LB patients (Stanek and Strle, 2018) and decorin is expressed in the nervous system (Hanemann et al., 1993; Kallmann et al., 2002), the aim was to study if the Dbp proteins are involved in colonization of the central nervous system. Our collaborators mutated specific lysines on the surface of the Bg SBK40 DbpA (K78/80/82A) and DbpB (K79A) proteins and performed binding experiments to study the effects on Bg decorin- and biglycan binding to human brain microvascular endothelial cells (HBMECs). To analyze the structural effects of the mutations, I created 3D structural models of the Bg SBK40 DbpA and DbpB proteins and used MD-simulations to study the stability of the proteins.

5.2.2 Structural overview of the Dbp proteins

Our experiments showed that the Dbp mutant proteins (DbpA K78/80/82A and DbpB K79A) were unable to bind to biglycan or decorin, and that *Borrelia* strains expressing the mutant proteins showed significantly reduced binding to HBMECs. To provide a structural explanation for this, I created homology models for the wt (wild type) and mutant DbpA and DbpB proteins. For DbpA, the Bg PBr DbpA structure (PDB ID: 2mtd (Morgan and Wang, 2015)), which has a sequence identity of 89.6% to DbpA, was used as a structural template for modeling (figure 15). For DbpB the *B. burgdorferi* sensu stricto (Bbss B31) DbpB structure (PDB ID: 2mvg, (Feng and Wang, 2015)), which has a sequence identity of 61 % to DbpB, was used. Visual inspection of the superimposed structures showed that the models were highly similar to the template structures, and this was also confirmed by the RMSD values between the template structures and the models which were 0.49 Å for DbpA and 0.51 Å for DbpB, indicating very similar structures. Furthermore, both models received very high scores in the ModFOLD (Maghrabi and Mcguffin, 2017) and ProQ (Wallner and Elofsson, 2003) quality evaluation. The ModFOLD global model quality score was above 0.5 for both DbpA and DbpB (score > 0.4 indicates complete and confident models) and the ProQ LG scores were 4.988 and 4.724 for DbpA and DbpB respectively (values >4 indicates extremely good models).

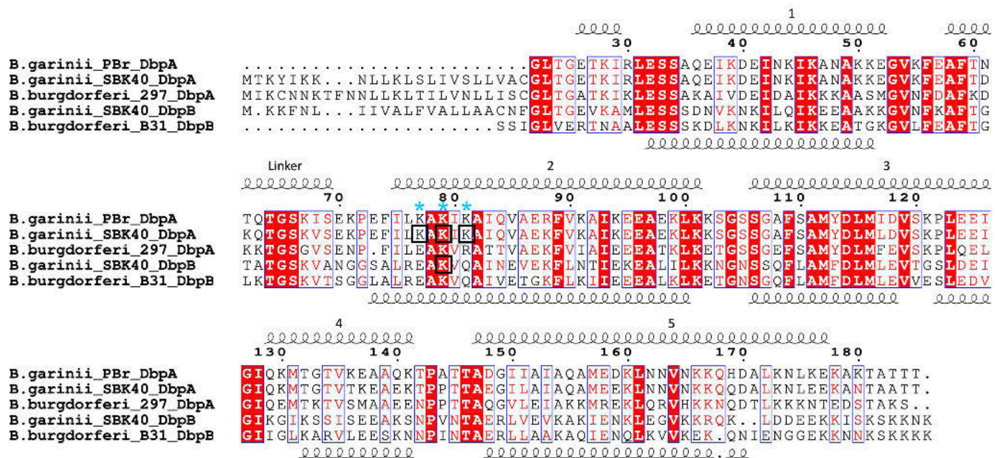


Figure 15. Multiple sequence alignment used for creating homology models for the Bg SBK40 DbpA and DbpB proteins. The structural templates Bg PBr DbpA (PDB ID: 2mtd, Morgan & Wang 2015) and Bbss B31 DbpB (PDB ID: 2mvg, (Feng & Wang 2015) are aligned with the Bg SBK40 DbpA and DbpB and with the extensively studied DbpA protein of Bbss 297. The secondary structure information for the templates is shown above (2mtd) and below (2mvg) the alignment. Black boxes and blue asterisks indicate the position of the mutated residues. Completely conserved residues are shown with a red background whereas similar residues are shown in red and boxed. Figure from Publication III.

The overall fold of the DbpA and DbpB models are similar to other known Dbp structures and consist of five alpha helices and a linker region connecting helices 1 and 2 (Feng and Wang, 2015; Fortune et al., 2014; Morgan and Wang, 2015; Wang, 2012). The linker forms an unstructured loop in DbpB and in most other Dbp structures, whereas in DbpA it has a well-structured helical shape. Helix 5 is significantly shorter in DbpB than in DbpA and the C-terminal end is consequently longer and more unstructured.

5.2.3 Sequence analysis of the *B. garinii* Dbp proteins

K80 in DbpA and K79 in DbpB correspond to K82 in *B. burgdorferi*, which is one of the three lysine residues (K82, K163 and K170) shown to be crucial for GAG-binding (Brown et al., 1999). From a multiple sequence alignment with all currently available Bg Dbp sequences (publication III, figure S1) it could be seen that this residue is completely conserved among all the included Bg sequences, highlighting its importance as a GAG-binding residue. The two additional lysines (K78, K82), found in close proximity to DbpA K80, correspond to K77 and K81 in Bg Pbr DbpA, where they have been shown to form a second binding site, independent from the canonical binding site (Morgan and Wang, 2015). Unlike K80, K78 and K82 are not conserved in all Bg sequences. K78 is mainly substituted for aspartate or glutamate whereas K82 is substituted for arginine or glutamine. In DbpB, the corresponding positions contain an aspartate (D77) and a glutamine (Q81), and thus these residues are unlikely to form a second binding site in the DbpB proteins. The substituted residues in the DbpA sequences are all large, polar or charged residues and the potential interactions formed with the GAG chains can possibly be maintained to some degree despite the differences. Arginine is positively charged just like lysine, and has been reported to form even stronger ionic interactions than lysine (Gandhi and Mancera, 2008). Glutamate, aspartate and glutamine are negatively charged residues; however, they could possibly form hydrogen bonds to the GAG chains. An alternative possibility is that the second binding site formed by these residues are only found in some of the Dbps, and in others are compensated for by residues in other positions. It has, in fact, been shown that some Dbps have several binding sites and that these can be uniquely adapted to each protein (Feng and Wang, 2015; Morgan et al., 2015; Morgan and Wang, 2015).

5.2.4 GAG-binding sites in DbpA and DbpB

The canonical GAG-binding site in DbpA is made up of residues K80, K161 and K168 (figure 16A). K78 and K82 are located on the opposite side of the protein where they form a second binding site, similar to Bg Pbr DbpA (Morgan and Wang, 2015). The second binding site in Bg Pbr DbpA consist of four residues (K77, K81, K173 and K176), is located on the opposite side of the protein in relation to the canonical binding site, and has been shown to act independently

(Morgan and Wang, 2015). In Bg SBK40 DbpA, K174 and E177 correspond to K173 and K176 in Bg Pbr DbpA, and K174 could participate in a second binding site while the negatively charged E177 is unable to form the ionic interactions normally found with GAG chains. The second binding site have the same affinity for GAGs as the canonical binding site (Morgan and Wang, 2015) and thus, by mutating residues from both binding sites we ensured that we obtained a non-functional protein that could be used to study the role of the decorin-binding proteins in adherence to HBMECs. In Bg Pbr DbpA, K173 and K176 were shown to have a greater impact on GAG-binding than K77 and K81. However, in our experiments the mutated DbpA (K87/80/82A), which still contains K174, did not bind to proteoglycans and the strain expressing the triple mutant only showed weak binding to HBMECs. This suggests that K174 does not play a significant role in GAG-binding in Bg SBK40 DbpA.

The canonical binding site in Bg SBK40 DbpB is made up of residues K79, K160 and R167 (figure 16B). Bbss B31 DbpB has been shown to contain two GAG-binding sites, and the second binding site is formed by the unstructured C-terminal tail which, interestingly, seems to be the primary GAG-binding site in Bbss B31 DbpB (Feng and Wang, 2015). In DbpB of Bg SBK40 however, the canonical binding site still plays an important role, as shown by the lack of proteoglycan binding seen in the DbpB K79A mutant. The C-terminal tail in Bbss B31 DbpB consists of four lysines whereas in Bg SBK40 DbpB it contains three lysines and an asparagine. Asparagine cannot form ionic interactions with the GAGs and this difference might explain the greater importance of the canonical binding site in Bg SBK40 DbpB. Furthermore, as the canonical binding site in Bbss B31 DbpB is made up of two lysines and a glutamine this suggests that the lysine-rich C-terminal tail compensates for the missing lysine in the canonical binding site. In contrast, the canonical GAG-binding residues in Bg SBK40 are two lysines and an arginine (K79, K160, R167) and as arginine have been shown to bind even more strongly than lysines, this further highlights the importance of the canonical GAG-binding site in Bg SBK40 DbpB. Further studies are needed to elucidate the role of the potential C-terminal binding site in Bg SBK40.

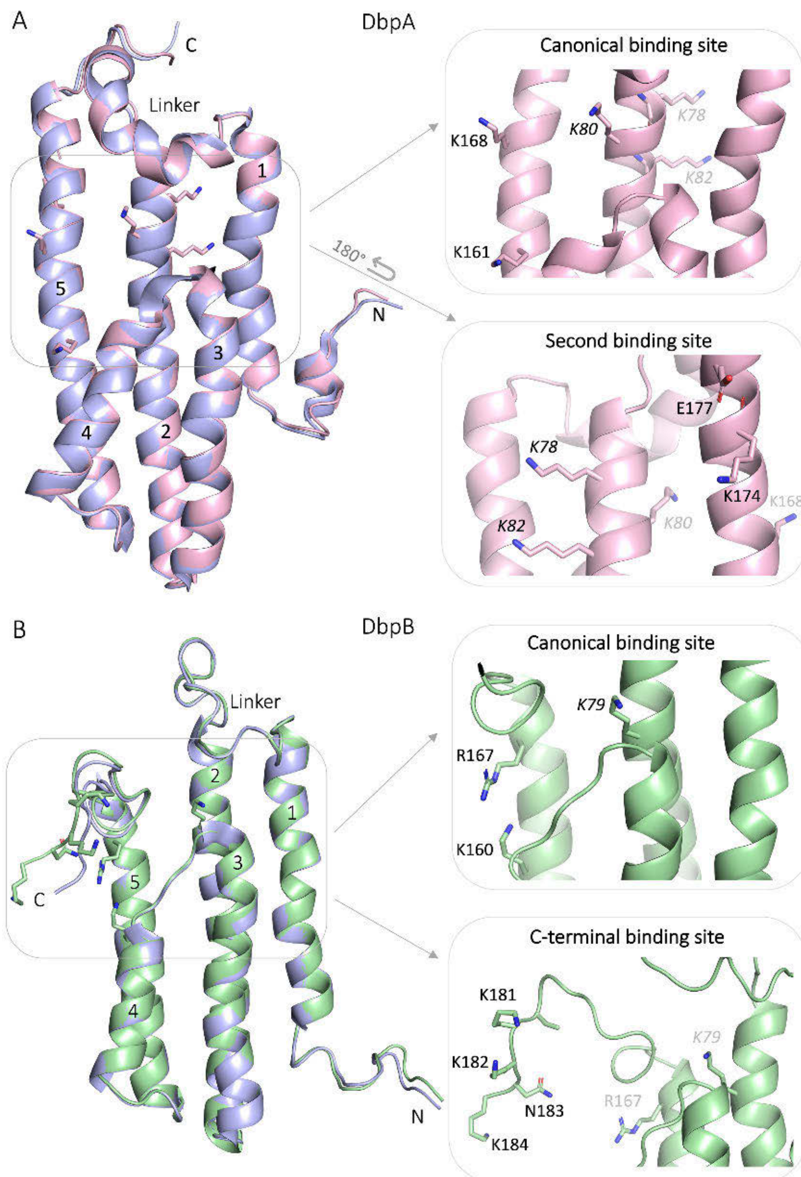


Figure 16. The GAG-binding sites in the 3D structures of Bg SBK40 DbpA and DbpB. A. Bg SBK40 DbpA (pink) superimposed on the structural template (light blue) Bg PBr DbpA (PDB ID: 2mtd, (Morgan & Wang, 2015)). The GAG-binding sites are shown in the close-ups. The canonical binding site consist of residues K80, K161 and K168. The second binding site is located on the opposite side of the protein and consist of residues K78 and K82 and residues K174 and E177. K174 and K177 corresponds to K173 and K176 in Bg PBr DbpA (see text for more details). B. Bg SBK40 DbpB (green) superimposed on the structural template (light blue) Bbss B31 DbpB (PDB ID: 2mvg, (Feng & Wang, 2015)). The canonical binding site of DbpB consist of residues K79, K160 and R167. The potential second binding site is located in the C-terminal tail (residues K181, K182, N183 and K184). The linker in DbpA forms a helical structure which is missing in DbpB. Helix 5 is shorter in DbpB than in DbpA and its C-terminal end therefore forms a longer flexible loop. Mutated residues are marked in italics and residues on the opposite side of the protein are in grey. Figure adapted from Publication III.

Since GAGs are known to interact with Dbps through electrostatic interactions, the effects of the mutations were further analyzed by studying the electrostatic surface potentials of the proteins. The mutations in the Dbp proteins cause significant changes in surface potential and since the Dbp proteins are known to bind to GAGs through ionic interactions, these changes can result in eliminated or reduced binding (figure 17). Electrostatic interactions are also important for long-range recognition of binding partners (Vascon et al., 2020), which means that changes in surface potential could make it less likely for the binding partners to form the initial contact needed for the interaction.

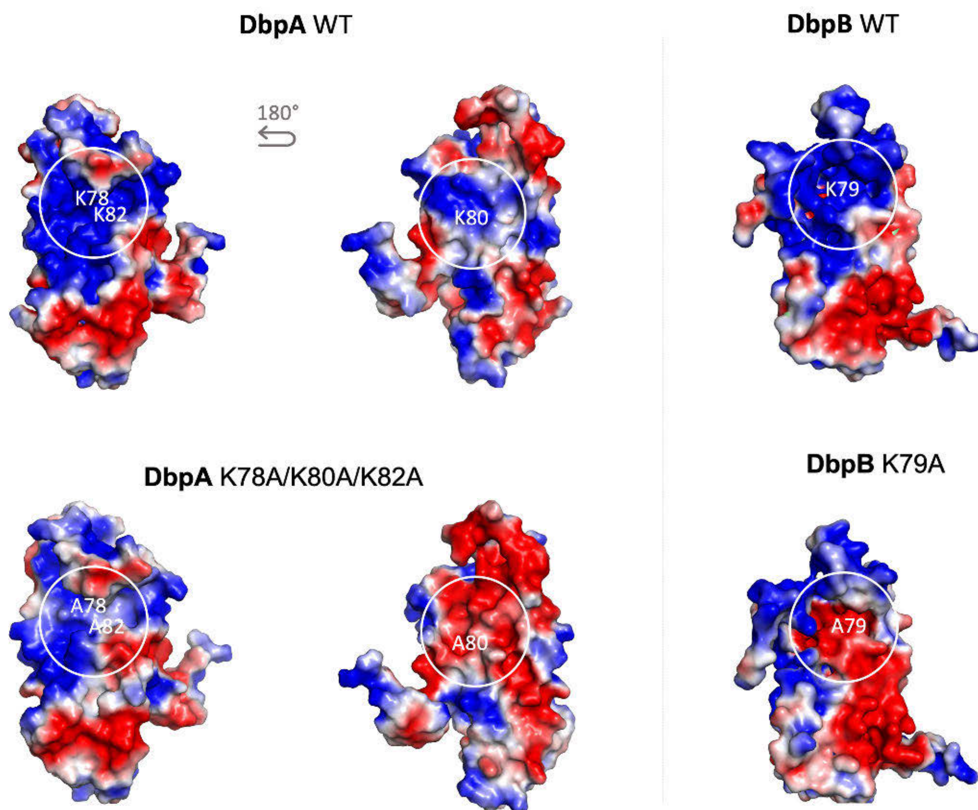


Figure 17. Electrostatic surface potential maps for the Bg SBK40 DbpA and DbpB models, calculated with the APBS tool (Adaptive Poisson–Boltzmann Solver) in PyMOL. The colors are in the range of -2 (red) to +2 (blue). Positively charged areas on the surface, created by lysine residues, are involved in ionic interactions with GAGs. The DbpA K78A/K80A/K82A mutations and the DbpB K79A mutation cause significant changes in the surface potential of the proteins, which prevents GAG-binding. Figure from Publication III.

5.2.5 Protein stability analysis

MD simulations were done to study whether the mutations would affect the stability of the protein. Experiments had shown that the mutant DbpA had a lower melting temperature than the wt protein, which could be the result of an unstable protein fold. However, the 3D fold of both the wt and the mutant proteins remained intact throughout the simulations and an increase in flexibility was only seen in the terminal regions and in the linker. Interestingly, the linker in wt DbpB was more flexible than in the mutant DbpB (Publication III, figure S3C) whereas no such difference could be observed for DbpA. This can be explained by the smaller size and hydrophobic nature of A79 compared to K79, which allows the linker to reach further down towards the canonical binding site. It has been shown that the shape of the linker has an impact on the GAG-binding affinity of DbpAs (Morgan et al., 2015), where a more compact shape gives a more exposed binding site and thus increases the binding affinity. As the linker in the mutant DbpB can extend further down towards the binding site it is possible that it would partly cover the binding site, preventing access to the canonical binding residues. The backbone residues near K79/A79 are also slightly more flexible in the mutant DbpB which can be explained by the different properties of lysine and alanine. K79 forms electrostatic interactions with the surrounding residues, which makes its positions more fixed, whereas A70 cannot form similar interactions and, thus, confers flexibility.

5.3 *Klebsiella pneumoniae* – VgrG4 and Sel1E

5.3.1 Introduction

The type VI secretion system (T6SS) is one of the many virulence factors used by bacteria to invade host cells. In *K. pneumoniae,* it has been shown to play a role in interbacterial competition, host invasion and colonization, as well as in type-1 fimbriae expression (Hsieh et al., 2019). The VgrG proteins are an essential part of the T6SS and bind to effector proteins, which they transport into the target cell, or carry internal effector domains (evolved VgrGs) (Hachani et al., 2014; Pukatzki et al., 2007). The VgrG proteins were first characterized in Rhs (recombination hotspot) elements in *E. coli* as proteins containing a recurring valine-glycine dipeptide repeat (Wang et al., 1998) and similar proteins have been found in a wide range of Gram-negative bacteria (Pukatzki et al., 2007). Our collaborators in Professor José A. Bengoechea's research group has shown that the VgrG4 protein in *K. pneumoniae* strain 52.145 (Kp 52.145) is toxic to other bacteria and to fungi, and that this effect is mediated through ROS generation (Storey et al., 2020). In contrast, VgrG1 and VgrG2 in the same *K. pneumoniae* strain did not have a cytotoxic effect.

VgrG proteins are often encoded in close proximity to an immunity protein within the same locus. The VgrG4 locus contains five genes, which code for sel-1 repeat proteins (Sel1A-E), a motif often involved in protein-protein interactions (D'Andrea and Regan, 2003). Two proteins with sel-1 repeats have been discovered also in *P. aeruginosa* (Yang et al., 2016) and these function as immunity proteins for the effector protein PldB, suggesting that the sel-1 proteins in the VgrG4 locus could have a similar function. Experiments performed by our collaborators subsequently demonstrated that Sel1E functions as the immunity protein for VgrG4 (Storey et al., 2020).

In this project, the functional differences between the three Kp 52.145 VgrG proteins were studied and the region of VgrG4 responsible for the observed cytotoxic effect was determined. A 3D structural model of VgrG4 was created to study the structural basis for its cytotoxic effect. The structure of the VgrG4 immunity protein Sel1E was also modeled, and compared to Sel1D, which does not inhibit the function of VgrG4, to identify residues potentially involved in binding to VgrG4.

5.3.2 *K. pneumoniae* VgrGs

The genome of Kp 52.145 comprises three T6SS loci, each coding for a VgrG protein: VgrG1 (W9BN84), VgrG2 (W9BP38/ W9BGC5) and VgrG4 (W9BAA4). VgrG proteins are trimeric proteins consisting of an N-terminal head part, where the three monomers reside side-by-side, and an elongated spike part, where the monomers are intertwined into a triangular-shaped beta helix (figure 18) (Cianfanelli et al., 2016). Many VgrGs also contain an extended C-terminal part (Records, 2011), which is often highly variable and plays a role in determining effector specificity (Bondage et al., 2016; Wettstadt et al., 2019). It also binds to PAAR proteins, which are shown to be vital for secretion of VgrG proteins (Wood et al., 2019b). A VgrG protein can require binding by a specific PAAR to function (Wood et al., 2019b) and PAAR proteins have also been shown to be able to deliver effectors to the VgrG protein by interacting with the effectors, and sometimes with chaperones, before mediating binding to the tip of the VgrG spike (Burkinshaw et al., 2018).

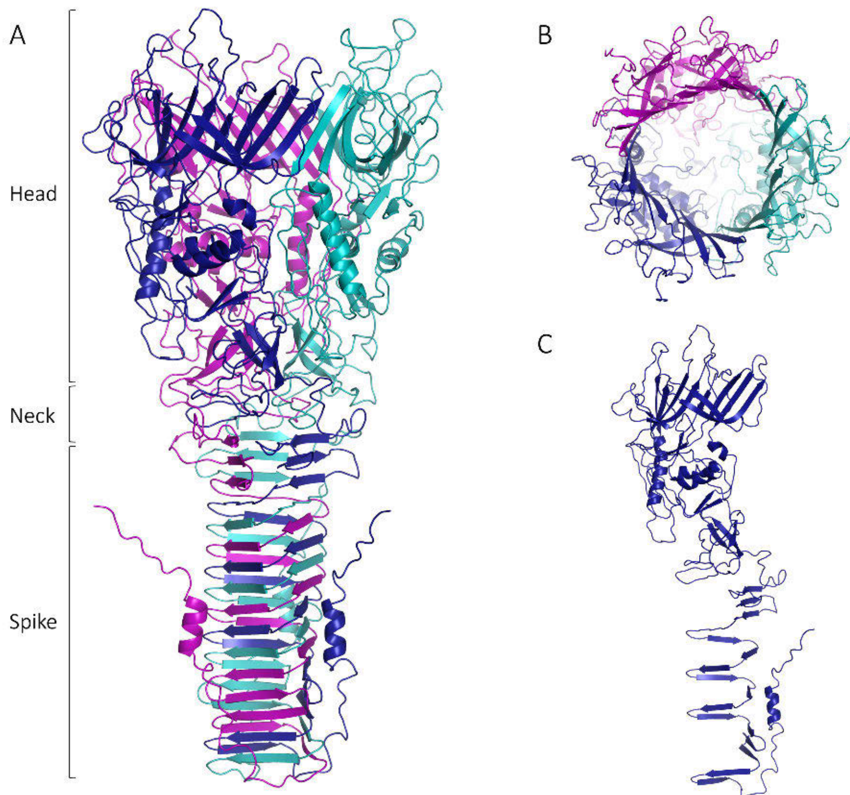


Figure 18. Generic structure of a VgrG protein, illustrated by the *P. aeruginosa* VgrG1 structure (PDB ID: 4uhv, (Spínola-Amilibia et al., 2016)). A. Trimeric structure of VgrG with each monomer colored in a separate color and the head, neck and spike regions depicted. The monomers in the spike part are intertwined to form the characteristic beta-helical fold. B. Top-down view of the head-part, showing the side-by-side arrangement of the three monomers in the head part. C. A single monomer of VgrG.

The Kp 52.145 VgrG proteins are evolved VgrGs and contain a conserved DUF2345 domain and an extended C-terminal part of varying length. VgrG1 consists of 851 amino acids, VgrG2 of 823 amino acids and VgrG4 of 899 amino acids, and they share about 40 % sequence identity. The differences between them are located in the C-terminal end of the proteins. As the experiments done by our collaborators showed that only VgrG4 causes a cytotoxic effect in target cells, this indicates that there is a clear functional difference between the Kp 52.145 VgrG proteins.

5.3.3 Structural characterization of VgrG4

The 3D model of VgrG4 was created in Modeller (Šali and Blundell, 1993) using the *Pseudomonas aeruginosa* VgrG1 (PaVgrG1) structure (PDB ID: 4uhv (Spínola-Amilibia et al., 2016) as the structural template. However, modeling was not straightforward as VgrG4 is about 250 amino acids longer than PaVgrG1 and has

a longer C-terminal part. PaVgrG1 has the typical trimeric structure of VgrGs, with an N-terminal head part and a C-terminal spike part, connected by a disorganized neck region (Spínola-Amilibia et al., 2016). Residues 1-563 (head, neck, N-terminal part of the spike) of VgrG4 could be confidently modeled based on the PaVgrG1 crystal structure (figure 19). The head part consists of 24 beta strands and 4 alpha helices from each monomer. The spike part of PaVgrG1 is made up of 16 beta strands that form the beta helical fold. In VgrG4 on the other hand, this repetitive beta strand fold is disrupted by an alpha helical part, which lacks a corresponding part in PaVgrG1. This part (residues 564-636) was modeled separately by I-TASSER as two long, almost parallel alpha helices that protrude from the spike fold. After the helices, the beta strand fold continues for 12 strands per monomer, as VgrG4 has 5 beta strands more than PaVgrG1. The VgrG spike structures are examples of beta helical folds which are based on a regular repeat pattern formed by eight residues: 5-6 residues make up the beta strand and 2-3 residues form a loop between two beta strands (Spínola-Amilibia et al., 2016). In each beta strand, every other residue point towards the interior of the spike, and these are predominantly hydrophobic, while the residues on the outside of the spike are mainly polar or charged residues.

The modeled part of VgrG4 ends with the putative PAAR interaction site, as the fold of the remaining part of the protein could not be modeled based on PaVgrG1 as the structural template. The VgrG4 sequence contains a region rich in prolines after the putative PAAR interaction site and as prolines are unfavored residues in beta strands, this indicates that the beta helical structure is most likely disrupted at this point and the remaining part of the protein forms a different kind of fold. In PaVgrG the C-terminal end turns upwards along the length of the spike where it forms alpha helices. As this leaves the end of the spike free to interact with PAAR proteins it is possible that the longer C-terminal part of VgrG4 could form even larger additional domains on the outside of the spike. This has indeed been seen in other VgrGs. The C-terminal ends of VgrG2b in *E. coli* (EAEC) Sci-1 and VgrG4b and VgrG5 in *P. aeruginosa* PAO1, contain TTR-like domains that can function as an adaptor for effector binding (Flaugnatti et al., 2016; Wettstadt et al., 2019). In addition to the TTR-like domain, VgrG2b in *P. aeruginosa* PAO1 also contains a metallopeptidase domain which causes toxic effects in target cells (Wood et al., 2019a). Likewise, the *V. cholerae* VgrG3 protein contains a peptidoglycan degrading domain at its C-terminal end (Brooks et al., 2013). This indicates that the large differences in length between VgrGs, and the variability of the C-terminal parts can correspond to very different functional domains. It also suggests that outside of the conserved head-spike structure common to all VgrGs, domains of varying size and function can exist.

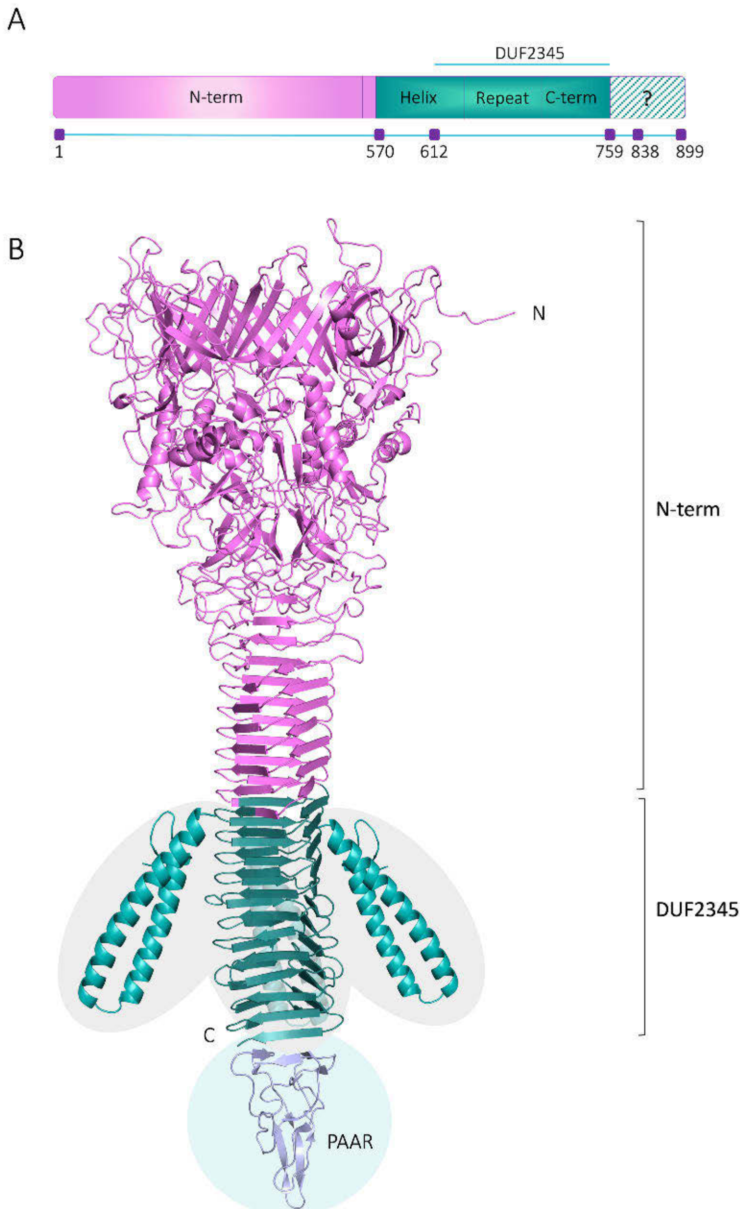


Figure 19. 3D model for *K. pneumoniae* VgrG4. A. The bar depicts the VgrG4 sequence and illustrates the different parts that were modeled: N-term: residues 1-563, Helix: residues 564-636, Repeat: residues 637-677, C-term: residues 678-739. The VgrG4 region (570-899) containing the DUF2345 domain (residues 612-759) can induce toxic effects in bacteria and yeast, while the 838-899 region cannot. B. Trimeric homology model of VgrG4. The teal-colored region depicts the functional part of VgrG4 (residues 570-899) and starts with the unique helical domains and continues with the beta helical spike, which is made up of 12 beta strands in each monomer. The independently modeled Helix part is inserted in the estimated location, however, its exact orientation in relation to the rest of the protein is not known. The 3D structure of *Vibrio cholerae* PAAR (PDB ID: 4jiv) is shown to indicate the location of the likely interaction site with the PAAR protein found in the same locus as VgrG4. Figure adapted from Publication IV.

Our collaborators performed experiments with truncated variants of VgrG4, which showed that residues 612-759 are sufficient for triggering the ROS-mediated toxic effect unique to VgrG4. These residues comprise a small part of the alpha helical part, as well as the remaining C-terminal beta strands (figure 19) and correspond to the predicted location of the DUF2345 domain. However, as the DUF2345 domain is also found in VgrG1 and VgrG2, the functional differences between the proteins are most likely explained by specific residue differences within this region, rather than any difference in their 3D fold.

After the publication of this work, a new structure of the spike part of an *E. coli* VgrG protein was published (Flaugnatti et al., 2020). This structure is significantly longer than the PaVgrG1 structure (figure 20) and contains two alpha helices located on the outside of the spike, just as predicted for *K. pneumoniae* VgrG4. The helices are of different lengths and are located parallel to the spike. Furthermore, the *E. coli* structure also contains a TTR (transthyretin-like) domain, which is located on the outside of the spike and connected to it by a short stretch of amino acids. However, the structure of this connecting region was not solved. The PaVgrG1 and the *E. coli* structures share the typical intertwined beta-helical repeat fold but differ slightly in the upper part of the spike, where the *E. coli* structure has a five-stranded antiparallel beta sheet, compared to the three-stranded antiparallel part in PaVgrG1. At the end of the beta-helical spike, the PaVgrG1 structure forms a loop that turns upwards along the length of the spike and ends in a short alpha helix, while the *E. coli* structure fold into the TTR domain. The *E. coli* VgrG protein is more similar in both sequence and length to VgrG4, compared to PaVgrG1, and the structure could therefore be used to refine the modeling of the VgrG4 spike part (unpublished). The residues involved in the helices are the same as was predicted by secondary structure predictions and overall, the fold of VgrG4 is highly similar to the *E. coli* structure.

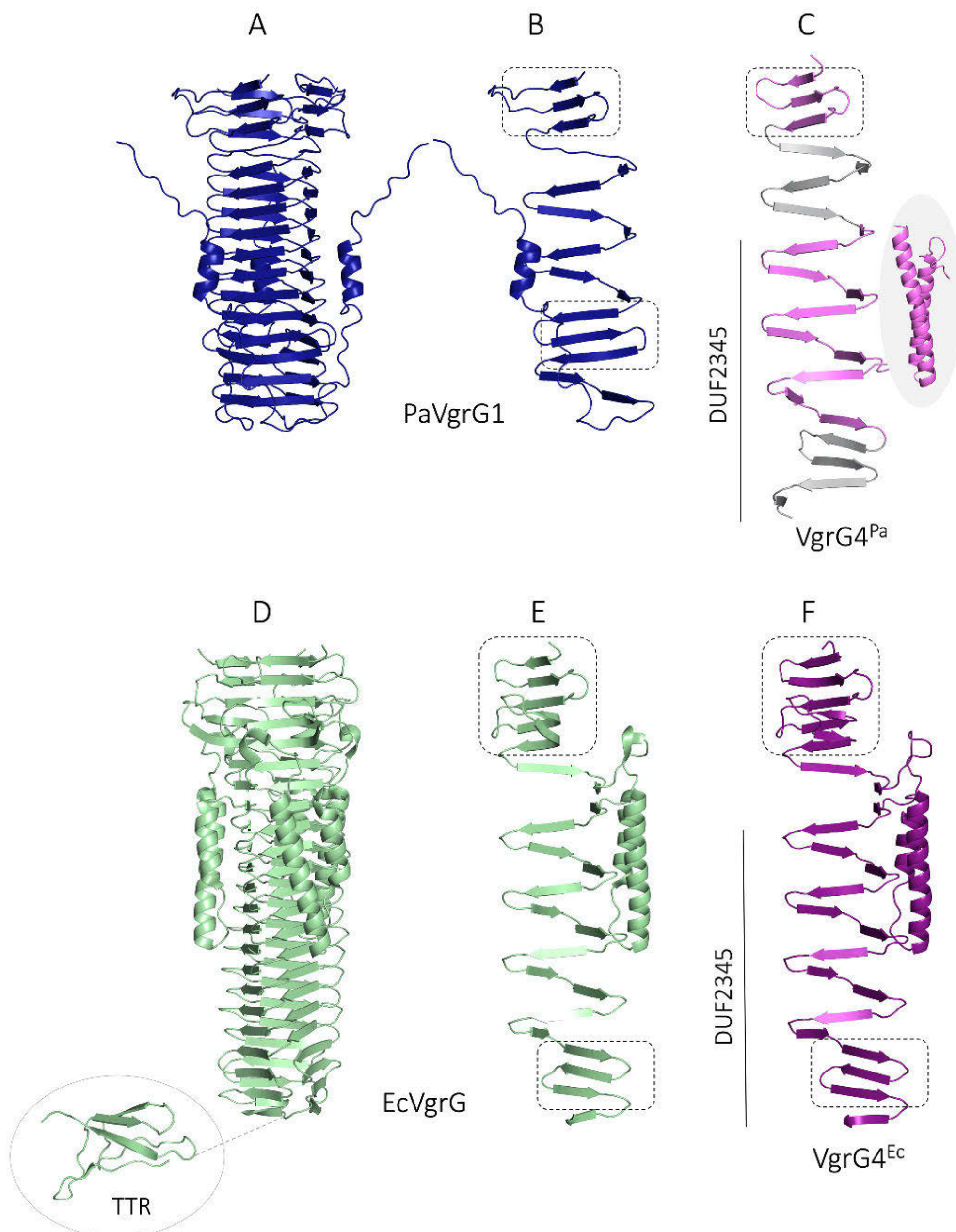


Figure 20. Comparison of the spike parts of the PaVgrG1*,* *E. coli* VgrG (EcVgrG), and the structural models for *K. pneumoniae* VgrG4. A-B: PaVgrG1 trimer and monomer. C: *K. pneumoniae* VgrG4 model based on PaVgrG1 as structural template (VgrG^{Pa}). D-E: *E. coli* VgrG1 trimer (with TTR domain) and monomer. F: *K. pneumoniae* VgrG4 model based on *E. coli* VgrG as structural template (VgrG4^{Ec}). The antiparallel beta sheets are shown in dashed boxes. The approximate location of the separately modeled helices is shown in grey background in C, and the parts of the model that were incorrect, in light of the refined model based on the *E. coli* structure, are shown in grey.

5.3.4 Sequence analysis and modelling of the immunity protein Sel1E

The Sel1 proteins are located in the same locus as VgrG4 and, according to experiments done by our collaborators, Sel1E is able to protect against VgrG4 toxicity, whereas Sel1D is not. To elucidate the structural basis for this functional difference, the structures of Sel1E and Sel1D were modeled and compared, and the evolutionary conservation of the sequences was studied.

The Sel1 proteins contain Sel1-like repeats (IPR006597) and belong to the Tetratricopeptide-like helical domain superfamily (IPR011990). The tetratricopeptide repeat (TPR) is an alpha/alpha repeat motif consisting of a consensus sequence of 34 amino acids that form two antiparallel alpha helices which are connected by a turn (Mittl and Schneider-Brachert, 2007). Alpha/alpha repeat proteins can contain a varying number of repeats and thus form proteins of different sizes. TPR motifs are mostly involved in protein-protein interactions and can be found in functionally very different proteins (D'Andrea and Regan, 2003). The Sel1 repeat (SLR) sequence is highly similar to the TPR motif but is longer (36-44 residues) (Mittl and Schneider-Brachert, 2007).

Sel1D and Sel1E share 69 % sequence identity and differences between them are mainly found in the terminal regions (figure 21). Sel1E has about 70 residues more than Sel1D and secondary structure predictions indicate that its C-terminal end contains a long loop region with very few predicted secondary structure elements. Sequence analyses did not find any information about potential domains or functions for this C-terminal region. Signal sequence predictions performed by SignalP (Almagro Armenteros et al., 2019) and LipoP (Juncker and Willenbrock, 2003) predicted an N-terminal lipoprotein signal sequence in Sel1D, but not in Sel1E. The N-terminus of Sel1D is slightly longer than in Sel1E and the first 18 residues are predominantly hydrophobic and ends with a cysteine, which is characteristic for all lipoprotein signal peptides in Gram-negative bacteria (Zückert, 2014). In Sel1E, the N-terminal is shorter and contains more polar or charged residues. Lipoprotein signal sequences target proteins for transport outside the cytoplasmic membrane, and the lack of signal sequence in Sel1E suggests that Sel1D and Sel1E can be located in different cellular compartments. However, ScanProsite (de Castro et al., 2006) and location predictions done by Psortdb (Peabody et al., 2016), CELLO (Yu et al., 2004) and Phobius (Käll et al., 2007) did not give any conclusive information about the cellular location of the proteins.

```
         1        10        20        30        40        50        60        70
Sel1D    MKGITTWIVLGSLLLSGCDQKSVHMEASSDTDPLSDISSSPAFTCQHETLPAPSADTDVLFKYARWLQKNNLLKQ
Sel1E    .........MACLLHSA...SSAKKDL..FVNPLSDTHAALAFTCAHQIIPEASADTDVLFKYARWLQKNNLLKQ

           80        90       100       110       120       130       140       150
Sel1D    DKSVDAQTERLYRIAAENGHYKASINLQNGALRWQFSLSSHEQFRLSQQLIAAGVATGYYLTAIYLERGVAGLQQ
Sel1E    DKSVDAQTERLYRIAAENGHYKASINLQNGALRGQFSLSSHEQFRLSQQLIAAGVATGYYLTAIYLERGVAGLQQ

              160       170       180       190       200       210       220
Sel1D    DPALALRYYRKAADEGNPQAQAYVGDKLAPVDRAPNIAHQMRRCAAEQGEGKAAAMLGINLQGKEDYQSAIEAFQ
Sel1E    DPALALRYYRKAADEGNPQAQAYVGDKLAPVDRAPDIAHQMRRCAAEQGEGKAAVMLGVNLQGKGYYRRAIEAFQ
                                            ●                  ●   ●     ●●  ●●

          230       240       250       260       270       280       290       300
Sel1D    LGVAAGNGSSARFLANGFSGPEPGDRLYYLAQQKDPERARRYKQIAKILSNYSYASPTVPEINDIVPLPPALLPE
Sel1E    LGVAAGDTSSALALSHGFDGPESSDELYYLAQQKDPERARRYKLITKILSNYSYASPTVPEINDIVPLPPAPLPE
               ●●   ●● ●●  ●                        ●

              310       320       330       340       350
Sel1D    WDGKLKWLEEREANVPPPKPSAALIEKLAKAKQLNPATGRPLPTSPDSGKYNRSAGQ..................
Sel1E    WDGKLKWLEEWEANIPPPAPDAALIEKLAQAKQLNPATGRPLPTSPDFEKDSVARLQCRSGEPCPQSGYWQPAWR

Sel1D    ....................................................
Sel1E    PREGMSEHAIRYFREGDIMPVEKVTFVRPRPWPLRDRLVVEAQETVWRRVSEA
```

Figure 21. Sequence alignment of Sel1D and Sel1E. The parts of the proteins that could be confidently modeled (Sel1D: residues 42-303, Sel1E: 28-289) are shown with pink background. The predicted signal sequence of Sel1D is shown with green background. Positions in Sel1D and Sel1E predicted by ConSurf to be both variable and exposed are marked with black dots. Residues conserved in both Sel1D and Sel1E are in bold.

For modeling of the Sel1 proteins, the threading-based method I-TASSER was used (Yang et al., 2015). We specified as a restraint for I-TASSER to use the crystal structure of *P. aeruginosa* Pa5087 (PDB ID: 5jkp, (Yang et al., 2016)) as structural template for modeling. Pa5087 shares 27 and 28 % sequence identity with Sel1D and Sel1E, respectively. However, since there was no suitable template available for the C-terminal parts of the proteins, these were modeled independently by I-TASSER. Residues 42-303 (Sel1D) and 28-289 (Sel1E) were modeled with high confidence based on Pa5087. The remaining C-terminal residues were modeled quite differently in Sel1D and Sel1E due to the lack of suitable template structures. The confidently modeled parts of Sel1D and Sel1E are highly similar to each other and to Pa5087, as shown by an RMSD value of 0.3 Å between both Sel1-proteins and Pa5087. The proteins consist of six SLR repeats, each made up of two antiparallel alpha helices (figure 22). In the MODFOLD (Maghrabi and Mcguffin, 2017) residue accuracy predictions the region between residues 42-303 (Sel1D) and 28-289 (Sel1E) shows a high accuracy, whereas the c-terminal parts show lower accuracy. Taken together, the I-TASSER models suggest that the N-terminal end of the Sel1D and Sel1E proteins share the typical fold of SLR-proteins while the structure of the C-terminal regions could not be accurately determined due to a lack of homologous structures.

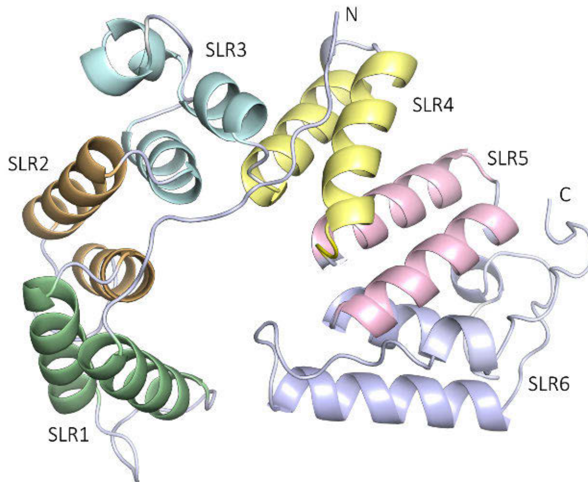


Figure 22. 3D structural model of *K. pneumoniae* Sel1E (residues 28-289). The six SLR-motifs are shown in different colors.

5.3.5 Prediction of potential binding interface between Sel1E and VgrG4

The structures of TPR- and SLR proteins are characterized by similar repeat motifs, but their overall structures can differ greatly due to a varying number of repeats, different angles between repeats, as well as the overall curvature of the protein (Mittl and Schneider-Brachert, 2007). Together, this gives the proteins a highly versatile structure, which can bind a wide range of molecules using several different binding modes (Perez-Riba and Itzhaki, 2019). The concave and convex surfaces created by the curved shape of most TPR- and SLR proteins give the proteins many potential binding sites.

The versatile SLR-structure, with its many possible binding sites, makes predicting the interaction site between VgrG4 and Sel1E a complex task. However, as there is a clear functional difference between Sel1D and Sel1E a detailed comparison of the structures and their surface properties made it possible to pinpoint residues that might play a role in interactions with other proteins. The ConSurf server (Ashkenazy et al., 2016) was used to depict evolutionary conservation based on multiple sequence alignments for the Sel1 proteins. ConSurf scores each residue based on a conservation scale, ranging from highly conserved to highly variable. The server also predicts whether a residue is buried or exposed, i.e., if the residue is located on the surface of the protein, and thus capable of interacting with other molecules, or if it is buried within the protein and more likely to have a structural role. When comparing the conservation patterns for Sel1D and Sel1E, virtually the same positions in both proteins are predicted as conserved and variable. Aside from the varying lengths of the C-terminal parts, there are only a few residue differences between Sel1D and Sel1E and most of these residues are also predicted by ConSurf to be exposed and highly variable (figure 23).

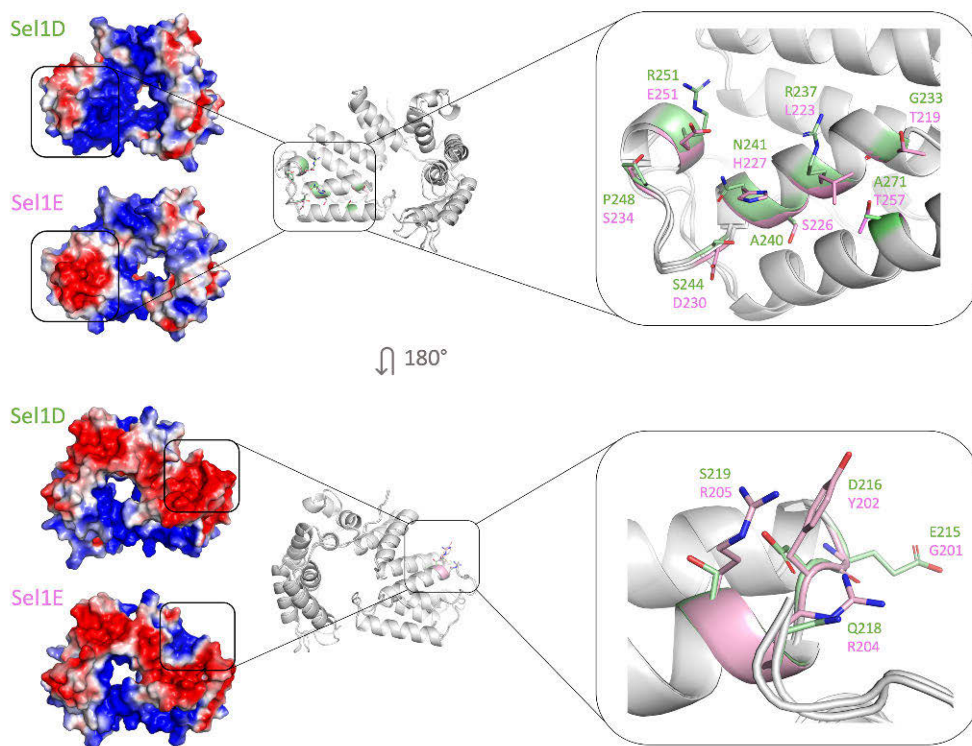


Figure 23. Comparison of electrostatic surface potentials for Sel1D and Sel1E. Due to the low confidence of the C-terminal parts of the models, only the high confidence regions of the models were used for the comparison. The close-ups show residues which contribute to the electrostatic potential differences, see text for more details. The electrostatic surfaces were calculated with the APBS tool (Adaptive Poisson–Boltzmann Solver) in PyMOL and the colour ranges from -2 to 2. Figure adapted from Publication IV.

In many TPR- and SLR structures, charged surface areas have been shown to play a role in ligand binding (Han et al., 2007; Zeytuni et al., 2011). The immunity proteins Pa5086, Pa5087 and Pa5088 in *P. aeruginosa* have significant differences in the electrostatic surface potentials (Wen et al., 2020; Yang et al., 2016). A similar pattern was found also in the Sel1 proteins when the electrostatic surface potential of the Sel1D and Sel1E structures was studied using the APBS plugin in PyMol. There were two regions with significant differences between SelD and Sel1E (figure 23), both located in the C-terminal half of the structures. In the first region, Sel1D has a more positively charged patch compared to Sel1E, which is explained mainly by the two arginines (figure 23A) found in this area (glutamate and leucine in Sel1E). In the other region, however, Sel1D has a glutamate and an aspartate which gives it a more negative charge than Sel1E, which has a glycine and a tyrosine (figure 23B). The residues in these two regions are among the residues that are predicted by ConSurf to be variable and surface exposed. Taken together this indicates that these residues

could be important for the function of the proteins and that their differing properties might enable them to bind to different partners. As the C-terminal part of VgrG4, which is responsible for its cytotoxic effect, consist of many charged surface residues, it is possible that Sel1E could form electrostatic interactions with specific residues on the surface of VgrG4 and thereby prevents its proper function.

6. Conclusion

In this thesis, virulence factors from *Borrelia* and *Klebsiella* have been studied and structure-function analyses have provided insights into their role in the virulence and survival mechanisms used by these pathogenic bacteria.

In publication I, 3D structural models were created for the *B. burgdorferi* Bmp proteins and their potential ligand-binding interactions were analyzed in detail. Furthermore, the evolutionary relationships between the *Borrelia* Bmp sequences were studied through a phylogenetic analysis. The structural analyses demonstrated that the Bmp proteins are substrate-binding proteins that transport purine nucleosides to ABC transporters for import into the bacteria and thus play a vital role in the purine salvage pathway. However, the ligand-binding site of BmpC differed from the other Bmp proteins to a significant degree, suggesting that it most likely has a different ligand preference than the other Bmp proteins. The phylogenetic analysis provided an improved classification of the Bmp proteins, showed that the previously identified BmpA2 proteins are more closely related to BmpD and that the ExP proteins most likely function as riboflavin-binding proteins. In publication II, the crystal structure of BmpD confirmed the structural fold seen in the models and showed that BmpD was bound to adenosine. Experimental results further demonstrated that BmpD can also bind to inosine, indicating that the Bmp proteins can possibly transport several different purine nucleosides. Further studies would be needed to fully determine the substrate scope of the Bmp proteins, and in particular that of BmpC.

In publication III, homology models were created for the *B. garinii* Dbp proteins and a detailed analysis of their potential GAG-binding sites were performed to explain the structural basis for the reduced GAG-binding seen in the lysine mutants. The analysis showed that the mutations cause a significant change in the electrostatic surface potential, thus reducing GAG-binding, which mainly occurs through ionic interactions. DbpA contains two GAG-binding sites, on opposite sides of the protein, whereas DbpB relies mainly on the canonical binding site, but potentially has a second binding site located at the C-terminal tail. MD-simulations were done to study the stability of the protein fold in the mutants, but this was not affected by the mutations and the proteins remained stable throughout the simulations. The Dbp proteins were shown to be important for *B. garinii* binding to HBMECs in the brain of mice, and further studies could be made to determine the exact binding mode of the Dbp proteins.

In publication IV, the 3D structures for the *K. pneumoniae* VgrG4 and Sel1E proteins were predicted, and their role in T6SS-mediated intermicrobial competitions and host interactions were studied. VgrG4 causes a ROS-mediated toxic effect in host cells and the effector function was located to the DUF2345 domain-containing region in the C-terminal of the protein. Sel1E was shown to

be the immunity protein of VgrG4 and was able to abolish the toxic effects induced by VgrG4. Structural analysis of Sel1E highlighted residues potentially involved in interactions with VgrG4. Future studies should focus on determining the full-length structure of VgrG4 and the exact binding mode of VgrG4 and Sel1E.

The proteins studied in this work illustrate two very different virulence strategies. The *Borrelia* Bmp proteins are used to obtain vital nutrients that the bacteria cannot synthesize themselves and the Dbp proteins ensure the tissue colonization required for a successful infection. While these methods can be seen as necessary for the survival of the bacteria, the VgrG4 protein of *K. pneumoniae*, on the other hand, is an example of a direct assault on a target cell, with the aim of causing damage. These examples demonstrate that interactions between a pathogen and its host, and between competing microbes, occur on many levels, from the direct attacks with harmful effectors to the more subtle evasion mechanism that keep the pathogen undetected by the immune system. Bacteria use several mechanisms simultaneously, usually in a highly coordinated manner, and in combination with the functional redundancy often seen among virulence factors, this makes studying these mechanisms far from trivial.

Gaining more knowledge of the complex network of interactions taking place between bacteria and its surroundings will aid our understanding of how bacteria avoid the immune system, how they are able to obtain nutrients and how they compete with other microorganisms to gain advantages. This is essential information needed when developing new treatment methods that are specifically targeted to cause maximum damage to vital bacterial pathways. Structural bioinformatics methods are a valuable tool for studying these bacterial virulence and survival mechanisms. The combined use of sequence-based studies with detailed structure analyses provides an in-depth knowledge of the molecular interactions behind these mechanisms, which is a requirement when developing drug molecules that can target specific proteins.

The result of this thesis contributes to our understanding of bacterial virulence and survival mechanisms in several ways. (i) By predicting the function of the Bmp proteins and improving the classification of the proteins (publication I), which greatly aids the study of functional differences between highly similar proteins. (ii) By providing a detailed structural explanation of ligand-binding properties (publication II and III), giving information on potential targets for inhibitory treatments. (iii) By predicting the possible interaction site between two proteins (publication IV), which provides valuable information for mutational studies into functionally important residues.

Despite the many kinds of antibiotics available today, bacterial infections are still a major cause of illness and death in the world. Only by understanding the details of the intricate mechanisms used by bacteria can we hope to develop efficient

treatments targeting the most vulnerable parts of the systems. Targeting bacterial virulence mechanisms could be an effective strategy, which would also potentially reduce the development of antibiotic resistance due to lowered selection pressures.

7. References

Aamer Mehmood, M., 2014. Use of Bioinformatics Tools in Different Spheres of Life Sciences. J. Data Mining Genomics Proteomics 05. https://doi.org/10.4172/2153-0602.1000158

Abu Kwaik, Y., Bumann, D., 2015. Host Delivery of Favorite Meals for Intracellular Pathogens. PLoS Pathog. 11, 1–8. https://doi.org/10.1371/journal.ppat.1004866

Actor, J.K., 2019. T Lymphocytes: Ringleaders of Adaptive Immune Function. Introd. Immunol. 45–62. https://doi.org/10.1016/b978-0-12-816572-0.00004-8

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., Charowhas, C., Clark, K., DiCuccio, M., Dondoshansky, I., Feolo, M., Funk, K., Geer, L.Y., Gorelenkov, V., Hlavina, W., Hoeppner, M., Holmes, B., Johnson, M., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Klimke, W., Krasnov, S., Kuznetsov, A., Landrum, M.J., Landsman, D., Lee, J.M., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Orris, R., Ostell, J., O'Sullivan, C., Palanigobu, V., Panchenko, A.R., Phan, L., Pruitt, K.D., Rodarmer, K., Rubinstein, W., Sayers, E.W., Schneider, V., Schoch, C.L., Schuler, G.D., Sherry, S.T., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Todorov, K., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., Wilbur, W.J., Yaschenko, E., Zbicz, K., 2017. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45, D12–D17. https://doi.org/10.1093/nar/gkw1071

Ajawatanawong, P., 2017. Molecular Phylogenetics: Concepts for a Newcomer. Adv. Biochem. Eng. Biotechnol. 160, 185–196. https://doi.org/10.1007/10_2016_49

Akira, S., Takeda, K., 2004. Toll-like receptor signalling. Nat. Rev. Immunol. 4, 499–511. https://doi.org/10.1038/nri1391

Akira, S., Uematsu, S., Takeuchi, O., 2006. Pathogen recognition and innate immunity. Cell 124, 783–801. https://doi.org/10.1016/j.cell.2006.02.015

Aldred, K.J., Kerns, R.J., Osheroff, N., 2014. Mechanism of Quinolone Action and Resistance. Biochemistry 53, 1565–1574.

Allen, J.P., Hauser, A.R., 2019. Diversity of Contact-Dependent Growth Inhibition Systems of Pseudomonas aeruginosa. J. Bacteriol. 201, 1–12. https://doi.org/10.1128/JB.00776-18

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37. https://doi.org/10.1038/s41587-019-0036-z

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Álvarez, D., Merino, S., Tomás, J.M., Benedí, V.J., Albertí, S., 2000. Capsular polysaccharide is a major complement resistance factor in lipopolysaccharide O side chain-deficient Klebsiella pneumoniae clinical isolates. Infect. Immun. 68, 953–955. https://doi.org/10.1128/IAI.68.2.953-955.2000

Aoki, S.K., Pamma, R., Hernday, A.D., Bickham, J.E., Braaten, B.A., Low, D.A., 2005. Microbiology: Contact-dependent inhibition of growth in Escherichia coli. Science (80-.). 309, 1245–1248. https://doi.org/10.1126/science.1115109

Aron, L., Alekshun, M., Perlee, L., Schwartz, I., Godfrey, H.P., Cabello, F.C., 1994. Cloning and DNA sequence analysis of bmpC, a gene encoding a potential membrane lipoprotein of Borrelia burgdorferi. FEMS Microbiol. Lett. 123, 75–82. https://doi.org/10.1111/j.1574-6968.1994.tb07204.x

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350. https://doi.org/10.1093/nar/gkw408

Auten, R.L., Davis, J.M., 2009. Oxygen toxicity and reactive oxygen species: The devil is in the details. Pediatr. Res. 66, 121–127. https://doi.org/10.1203/PDR.0b013e3181a9eafb

Barcena-Uribarri, I., Thein, M., Sacher, A., Bunikis, I., Bonde, M., Bergström, S., Benz, R., 2010. P66 porins are present in both Lyme disease and relapsing fever spirochetes: a comparison of the biophysical properties of P66 porins from six Borrelia species. Biochim Biophys Acta. 1798, 1197–1203.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L.G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Zhang, J., 2017. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169. https://doi.org/10.1093/nar/gkw1099

Bengoechea, J.A., Sa Pessoa, J., 2019. Klebsiella pneumoniae infection biology: Living to counteract host defences. FEMS Microbiol. Rev. 43, 123–144. https://doi.org/10.1093/femsre/fuy043

Benkert, P., Künzli, M., Schwede, T., 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res. 37, 510–514. https://doi.org/10.1093/nar/gkp322

Benoit, V.M., Fischer, J.R., Lin, Y.P., Parveen, N., Leong, J.M., 2011. Allelic variation of the Lyme disease spirochete adhesin DbpA influences spirochetal binding to decorin, dermatan sulfate, and mammalian cells. Infect. Immun. 79, 3501–3509. https://doi.org/10.1128/IAI.00163-11

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res. 28, 235–42.

Bhoite, S., Gerven, N. Van, Chapman, M.R., Remaut, H., Brussels, B., Brussel, V.U., Biology, S., 2020. Curli Biogenesis: Bacterial amyloid assembly by the Type VIII secretion pathway 8, 1–12. https://doi.org/10.1128/ecosalplus.

ESP-0037-2018.Curli
Bliven, K.A., Maurelli, A.T., 2016. Evolution of Bacterial Pathogens Within the Human Host. Microbiol. Spectr. 4, 1–13. https://doi.org/10.1128/microbiolspec.vmbf-0017-2015
Bondage, D.D., Lin, J.-S., Ma, L.-S., Kuo, C.-H., Lai, E.-M., 2016. VgrG C terminus confers the type VI effector transport specificity and is required for binding with PAAR and adaptor-effector complex. Proc. Natl. Acad. Sci. U. S. A. 113, E3931-40. https://doi.org/10.1073/pnas.1600428113
Bonilla, F.A., Oettgen, H.C., 2010. Adaptive immunity. J. Allergy Clin. Immunol. 125, S33–S40. https://doi.org/10.1016/j.jaci.2009.09.017
Brestoff, J., Artis, D., 2013. Commensal bacteria at the interface of host metabolism and the immune systemPopulation genetic tools to dissect innate immunity in humans. Nat Immunol 14, 676–684. https://doi.org/10.1038/ni.2640.Commensal
Brissette, C.A., Gaultney, R.A., 2014. That's my story, and I'm sticking to it-an update on B. burgdorferi adhesins. Front. Cell. Infect. Microbiol. 4, 1–13. https://doi.org/10.3389/fcimb.2014.00041
Brooks, T.M., Unterweger, D., Bachmann, V., Kostiuk, B., Pukatzki, S., 2013. Lytic activity of the Vibrio cholerae type VI secretion toxin VgrG-3 is inhibited by the antitoxin TsaB. J. Biol. Chem. https://doi.org/10.1074/jbc.M112.436725
Brown, D., Sjölander, K., 2006. Functional classification using phylogenomic inference. PLoS Comput. Biol. 2, 0479–0483. https://doi.org/10.1371/journal.pcbi.0020077
Brown, E.L., Guo, B.P., O'Neal, P., Höök, M., 1999. Adherence of Borrelia burgdorferi . J. Biol. Chem. 274, 26272–26278. https://doi.org/10.1074/jbc.274.37.26272
Brown, S.A., Palmer, K.L., Whiteley, M., 2008. Revisiting the host as a growth medium. Nat. Rev. Microbiol. 6, 657–666. https://doi.org/10.1038/nrmicro1955
Buddelmeijer, N., 2015. The molecular mechanism of bacterial lipoprotein modification-How, when and why? FEMS Microbiol. Rev. 39, 246–261. https://doi.org/10.1093/femsre/fuu006
Burkinshaw, B.J., Liang, X., Wong, M., Le, A.N.H., Lam, L., Dong, T.G., 2018. A type VI secretion system effector delivery mechanism dependent on PAAR and a chaperone-co-chaperone complex. Nat. Microbiol. 3, 632–640. https://doi.org/10.1038/s41564-018-0144-4
Bush, K., 2012. Antimicrobial agents targeting bacterial cell walls and cell membranes. OIE Rev. Sci. Tech. 31, 43–56. https://doi.org/10.20506/rst.31.1.2096
Cabeen, M.T., Jacobs-Wagner, C., 2005. Bacterial cell shape. Nat. Rev. Microbiol. 3, 601–610. https://doi.org/10.1038/nrmicro1205
Casadevall, A., Pirofski, L., 2019. Benefits and Costs of Animal Virulence for Microbes 1–7.
Cassat, J.E., Skaar, E.P., 2013. Iron in infection and immunity. Cell Host Microbe 13, 509–519. https://doi.org/10.1016/j.chom.2013.04.010
Chandra, N., Anand, P., Yeturu, K., 2010. Structural bioinformatics: Deriving biological insights from protein structures. Interdiscip. Sci. Comput. Life Sci. 2, 347–366. https://doi.org/10.1007/s12539-010-0045-6
Chen, C., Huang, H., Wu, C., 2017. Protein Bioinformatics Databases and Resources. Methods Mol. Biol. 1558, 3–39.
Cherrak, Y., Flaugnatti, N., Durand, E., Journet, L., Cascales, E., 2019. Structure and Activity of the Type VI Secretion System. Protein Secret. Bact. 329–342. https://doi.org/10.1128/microbiolspec.psib-0031-2019
Chowdhury, B., Garai, G., 2017. A review on

multiple sequence alignment from the perspective of genetic algorithm. Genomics 109, 419–431. https://doi.org/10.1016/j.ygeno.2017.06.007
Cianfanelli, F.R., Monlezun, L., Coulthurst, S.J., 2016. Aim, Load, Fire: The Type VI Secretion System, a Bacterial Nanoweapon. Trends Microbiol. 24, 51–62. https://doi.org/10.1016/j.tim.2015.10.005
Cornelissen, C.N., 2018. Subversion of nutritional immunity by the pathogenic Neisseriae. Pathog. Dis. 76, 1–14. https://doi.org/10.1093/femspd/ftx112
Cortés, G., Borrell, N., De Astorza, B., Gómez, C., Sauleda, J., Albertí, S., 2002. Molecular analysis of the contribution of the capsular polysaccharide and the lipopolysaccharide O side chain to the virulence of Klebsiella pneumoniae in a murine model of pneumonia. Infect. Immun. 70, 2583–2590. https://doi.org/10.1128/IAI.70.5.2583-2590.2002
Coulthurst, S., 2019. The Type VI secretion system: A versatile bacterial weapon. Microbiol. (United Kingdom) 165, 503–515. https://doi.org/10.1099/mic.0.000789
D'Andrea, L.D., Regan, L., 2003. TPR proteins: The versatile helix. Trends Biochem. Sci. 28, 655–662. https://doi.org/10.1016/j.tibs.2003.10.007
Davidson, A.L., Dassa, E., Orelle, C., Chen, J., 2008a. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol. Mol. Bio. Rev. 72, 317–364, table of contents. https://doi.org/10.1128/MMBR.00031-07
Davidson, A.L., Dassa, E., Orelle, C., Chen, J., 2008b. Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. Microbiol. Mol. Biol. Rev. 72, 317–364. https://doi.org/10.1128/MMBR.00031-07
de Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34. https://doi.org/10.1093/nar/gkl124
Deka, R.K., Brautigam, C.A., Biddy, B.A., Liu, W.Z., Norgard, M. V., 2013. Evidence for an ABC-type riboflavin transporter system in pathogenic spirochetes. MBio. https://doi.org/10.1128/mBio.00615-12
Deka, R.K., Brautigam, C.A., Yang, X.F., Blevins, J.S., Machius, M., Tomchick, D.R., Norgard, M. V., 2006. The PnrA (Tp0319; TmpC) lipoprotein represents a new family of bacterial purine nucleoside receptor encoded within an ATP-binding cassette (ABC)-like operon in Treponema pallidum. J. Biol. Chem. https://doi.org/10.1074/jbc.M511405200
Dey, S., Chakravarty, A., Guha Biswas, P., De Guzman, R.N., 2019. The type III secretion system needle, tip, and translocon. Protein Sci. 28, 1582–1593. https://doi.org/10.1002/pro.3682
Dobrikova, E.Y., Bugrysheva, J., Cabello, F.C., 2001. Two independent transcriptional units control the complex and simultaneous expression of the bmp paralogous chromosomal gene family in Borrelia burgdorferi. Mol. Microbiol. 39, 370–378. https://doi.org/10.1046/j.1365-2958.2001.02220.x
Dong, T.G., Dong, S., Catalano, C., Moore, R., Liang, X., Mekalanos, J.J., 2015. Generation of reactive oxygen species by lethal attacks from competing microbes. Proc. Natl. Acad. Sci. U. S. A. 112, 2181–2186. https://doi.org/10.1073/pnas.1425007112
Dönhöfer, A., Franckenberg, S., Wickles, S., Berninghausen, O., Beckmann, R., Wilson, D.N., 2012. Structural basis for TetM-mediated tetracycline resistance. Proc. Natl. Acad. Sci. U. S. A. 109, 16900–16905. https://doi.org/10.1073/pnas.1208037109

Dorn, M., E Silva, M.B., Buriol, L.S., Lamb, L.C., 2014. Three-dimensional protein structure prediction: Methods and computational strategies. Comput. Biol. Chem. 53, 251–276. https://doi.org/10.1016/j.compbiolchem.2014.10.001
Dowdell, A.S., Murphy, M.D., Azodi, C., Swanson, S.K., Florens, L., Chen, S., Zückert, W.R., 2017. Comprehensive spatial analysis of the Borrelia burgdorferi lipoproteome reveals a compartmentalization bias toward the bacterial surface. J. Bacteriol. 199, 1–20. https://doi.org/10.1128/JB.00658-16
Duesbery, N.S., Webb, C.P., Leppla, S.H., Gordon, V.M., Klimpel, K.R., Copeland, T.D., Ahn, N.G., Oskarsson, M.K., Fukasawa, K., Paull, K.D., Vande Woude, G.F., 1998. Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor. Science (80-.). 280, 734–737. https://doi.org/10.1126/science.280.5364.734
Dunkelberger, J.R., Song, W.C., 2010. Complement and its role in innate and adaptive immune responses. Cell Res. 20, 34–50. https://doi.org/10.1038/cr.2009.139
Dyson, H.J., Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208. https://doi.org/10.1038/nrm1589
Ernst, C.M., Staubitz, P., Mishra, N.N., Yang, S.J., Hornig, G., Kalbacher, H., Bayer, A.S., Kraus, D., Peschel, A., 2009. The bacterial defensin resistance protein MprF consists of separable domains for lipid lysinylation and antimicrobial peptide repulsion. PLoS Pathog. 5, 1–9. https://doi.org/10.1371/journal.ppat.1000660
Eswarappa, S.M., Panguluri, K.K., Hensel, M., Chakravortty, D., 2008. The yejABEF operon of Salmonella confers resistance to antimicrobial peptides and contributes to its virulence. Microbiology 154, 666–678. https://doi.org/10.1099/mic.0.2007/011114-0
Fan, E., Chauhan, N., Udatha, D.B.R.K.G., Leo, J.C., Linke, D., 2016. Type V Secretion Systems in Bacteria. Microbiol. Spectr. 4, 1–24. https://doi.org/10.1128/microbiolspec.vmbf-0009-2015
Felsenstein, J., 1985. Confidence Limits on Phylogenies : An Approach Using the Bootstrap Author (s): Joseph Felsenstein Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2408678. Evolution (N. Y). 39, 783–791.
Feng, W., Wang, X., 2015. Structure of decorin binding protein B from Borrelia burgdorferi and its interactions with glycosaminoglycans. Biochim. Biophys. Acta - Proteins Proteomics 1854, 1823–1832. https://doi.org/10.1016/j.bbapap.2015.08.003
Fernández-Villa, D., Aguilar, M.R., Rojo, L., 2019. Folic acid antagonists: Antimicrobial and immunomodulating mechanisms and applications. Int. J. Mol. Sci. 20, 1–30. https://doi.org/10.3390/ijms20204996
Finlay, B.B., McFadden, G., 2006. Anti-immunology: Evasion of the host immune system by bacterial and viral pathogens. Cell 124, 767–782. https://doi.org/10.1016/j.cell.2006.01.034
Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y., Dosztanyi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D.A., Necci, M., Nuka, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C., Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Xenarios, I., Yeh, L.S., Young, S.Y., Mitchell, A.L., 2017. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 45, D190–D199. https://doi.org/10.1093/nar/gkw1107

Fischer, J.R., Parveen, N., Magoun, L., Leong, J.M., 2003. Decorin-binding proteins A and B confer distinct mammalian cell type-specific attachment by Borrelia burgdorferi, the Lyme disease spirochete. Proc. Natl. Acad. Sci. U. S. A. 100, 7307–7312. https://doi.org/10.1073/pnas.1231043100

Flaugnatti, N., Le, T.T.H., Canaan, S., Aschtgen, M.S., Nguyen, V.S., Blangy, S., Kellenberger, C., Roussel, A., Cambillau, C., Cascales, E., Journet, L., 2016. A phospholipase A1 antibacterial Type VI secretion effector interacts directly with the C-terminal domain of the VgrG spike protein for delivery. Mol. Microbiol. 99, 1099–1118. https://doi.org/10.1111/mmi.13292

Flaugnatti, N., Rapisarda, C., Rey, M., Beauvois, S.G., Nguyen, V.A., Canaan, S., Durand, E., Chamot-Rooke, J., Cascales, E., Fronzes, R., Journet, L., 2020. Structural basis for loading and inhibition of a bacterial T6 SS phospholipase effector by the VgrG spike . EMBO J. 39, 1–14. https://doi.org/10.15252/embj.2019104129

Forthal, D.N., 2014. Functions of antibodies. Microbiol spectr 2, 1–17.

Fortune, D.E., Lin, Y.P., Deka, R.K., Groshong, A.M., Moore, B.P., Hagman, K.E., Leong, J.M., Tomchick, D.R., Blevins, J.S., 2014. Identification of lysine residues in the Borrelia burgdorferi DbpA adhesin required for murine infection. Infect. Immun. 82, 3186–3198. https://doi.org/10.1128/IAI.02036-14

Frain, K.M., Robinson, C., van Dijl, J.M., 2019. Transport of Folded Proteins by the Tat System. Protein J. 38, 377–388. https://doi.org/10.1007/s10930-019-09859-y

França, T.C.C., 2015. Homology modeling: an important tool for the drug discovery. J. Biomol. Struct. Dyn. 33, 1780–1793. https://doi.org/10.1080/07391102.2014.971429

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., Van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O., Venter, J.C., 1997. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature 390, 580–586. https://doi.org/10.1038/37551

Galán, J.E., 2009. Common themes in the design and function of bacterial effectors. Cell Host Microbe 5, 571–579.

Gandhi, N.S., Mancera, R.L., 2008. The structure of glycosaminoglycans and their interactions with proteins. Chem. Biol. Drug Des. 72, 455–482. https://doi.org/10.1111/j.1747-0285.2008.00741.x

García-Bayona, L., Comstock, L.E., 2018. Bacterial antagonism in host-associated microbial communities. Science (80-.). 361. https://doi.org/10.1126/science.aat2456

García, B., Merayo-Lloves, J., Martin, C., Alcalde, I., Quirós, L.M., Vazquez, F., 2016. Surface proteoglycans as mediators in bacterial pathogens infections. Front. Microbiol. 7, 1–11. https://doi.org/10.3389/fmicb.2016.00220

Gauthier, J., Vincent, A.T., Charette, S.J., Derome, N., 2019. A brief history of bioinformatics. Brief. Bioinform. 20.

Gherardini, F., Boylan, J., Lawrence, K., Skare, J., 2010. Metabolism and physiology of Borrelia, in: Samuels, S., Radolf, J.D. (Eds.), Borrelia: Molecular Biology, Host Interaction and Pathogenesis. Caister Academic Press, Norfolk, United Kingdom., pp. 103–138.

Granato, E.T., Meiller-Legrand, T.A., Foster, K.R., 2019. The Evolution and Ecology of Bacterial Warfare. Curr. Biol. 29, R521–R537. https://doi.org/10.1016/j.cub.2019.04.024

Green, E.R., Mecsas, J., 2016. Bacterial

Secretion Systems: An Overview. Microbiol. Spectr. 4, 1–19. https://doi.org/10.1128/microbiolspec.vmbf-0012-2015
Green, E.R., Mecsas, J., 2015. Bacterial Secretion Systems – An overview. Am. Soc. Microbiol. 4, 1–32. https://doi.org/10.1128/microbiolspec.VMBF-0012-2015.Bacterial
Grigg, J.C., Ukpabi, G., Gaudin, C.F.M., Murphy, M.E.P., 2010. Structural biology of heme binding in the Staphylococcus aureus Isd system. J. Inorg. Biochem. 104, 341–348. https://doi.org/10.1016/j.jinorgbio.2009.09.012
Guo, B.P., Brown, E.L., Dorward, D.W., Rosenberg, L.C., Höök, M., 1998. Decorin-binding adhesins from Borrelia burgdorferi. Mol. Microbiol. 30, 711–723. https://doi.org/10.1046/j.1365-2958.1998.01103.x
Guo, B.P., Norris, S.J., Rosenberg, L.C., Hook, M., 1995. Adherence of Borrelia burgdorferi to the proteoglycan decorin. Infect. Immun. 63, 3467–3472. https://doi.org/10.1128/iai.63.9.3467-3472.1995
Hachani, A., Allsopp, L.P., Oduko, Y., Filloux, A., 2014. The VgrG proteins are "à la carte" delivery systems for bacterial type VI effectors. J. Biol. Chem. 289, 17872–17884. https://doi.org/10.1074/jbc.M114.563429
Haddad, Y., Adam, V., Heger, Z., 2020. Ten quick tips for homology modeling of high-resolution protein 3D structures. PLoS Comput. Biol. 16, 1–19. https://doi.org/10.1371/journal.pcbi.1007449
Hajam, I.A., Dar, P.A., Shahnawaz, I., Jaume, J.C., Lee, J.H., 2017. Bacterial flagellin—a potent immunomodulatory agent. Exp. Mol. Med. 49, e373–e373. https://doi.org/10.1038/emm.2017.172
Han, D., Oh, J., Kim, K., Lim, H., Kim, Y., 2007. Crystal structure of YrrB: A TPR protein with an unusual peptide-binding site. Biochem. Biophys. Res. Commun. 360, 784–790. https://doi.org/10.1016/j.bbrc.2007.06.129
Han, Y., Wang, T., Chen, G., Pu, Q., Liu, Q., Zhang, Y., Xu, L., Wu, M., Liang, H., 2019. A Pseudomonas aeruginosa type VI secretion system regulated by CueR facilitates copper acquisition. PLoS Pathog. 15, 1–25. https://doi.org/10.1371/journal.ppat.1008198
Hanemann, C.O., Kuhn, G., Lie, A., Gillen, C., Bosse, F., Spreyer, P., Muller, H.W., 1993. Expression of decorin mRNA in the nervous system of rat. J. Histochem. Cytochem. 41, 1383–1391. https://doi.org/10.1177/41.9.8354878
Hanson, M.S., Cassatt, D.R., Guo, B.P., Patel, N.K., Mccarthy, M.P., Dorward, D.W., Höök, M., 1998. Active and passive immunity against Borrelia burgdorferi decorin binding protein A (DbpA) protects against infection. Infect. Immun. 66, 2143–2153. https://doi.org/10.1128/iai.66.5.2143-2153.1998
Hassan, M., Kjos, M., Nes, I.F., Diep, D.B., Lotfipour, F., 2012. Natural antimicrobial peptides from bacteria: Characteristics and potential applications to fight against antibiotic resistance. J. Appl. Microbiol. 113, 723–736. https://doi.org/10.1111/j.1365-2672.2012.05338.x
Heesterbeek, D.A.C., Angelier, M.L., Harrison, R.A., Rooijakkers, S.H.M., 2018. Complement and Bacterial Infections: From Molecular Mechanisms to Therapeutic Applications. J. Innate Immun. 10, 455–464. https://doi.org/10.1159/000491439
Hennigar, S.R., McClung, J.P., 2014. Nutritional Immunity: Starving Pathogens of Trace Minerals. Am. J. Lifestyle Med. 10, 170–173. https://doi.org/10.1177/1559827616629117
Henzler-Wildman, K., Kern, D., 2007. Dynamic personalities of proteins. Nature 450, 964–972. https://doi.org/10.1038/nature06522
Hersch, S.J., Manera, K., Dong, T.G., 2020.

Defending against the Type Six Secretion System: beyond Immunity Genes. Cell Rep. 33, 108259. https://doi.org/10.1016/j.celrep.2020.108259
Hibbing, M.E., Fuqua, C., Parsek, M.R., Peterson, S.B., 2010. Bacterial competition: Surviving and thriving in the microbial jungle. Nat. Rev. Microbiol. 8, 15–25. https://doi.org/10.1038/nrmicro2259
Hileman, R.E., Fromm, J.R., Weiler, J.M., Linhardt, R.J., 1998. Glycosaminoglycan-protein interactions: Definition of consensus sites in glycosaminoglycan binding proteins. BioEssays 20, 156–167. https://doi.org/10.1002/(SICI)1521-1878(199802)20:2<156::AID-BIES8>3.0.CO;2-R
Hillis, D.M., Bull, J.J., 2010. Society of Systematic Biologists An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis Published by : Oxford University Press for the Society of Systematic Biologists Stable URL : http://www.jstor.org/stable/299242, 182–192.
Hollingsworth, S.A., Dror, R.O., 2018. Molecular Dynamics Simulation for All. Neuron 99, 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011
Holmes, A.H., Moore, L.S.P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P.J., Piddock, L.J.V., 2016. Understanding the mechanisms and drivers of antimicrobial resistance. Lancet 387, 176–187. https://doi.org/10.1016/S0140-6736(15)00473-0
Hood, R.D., Singh, P., Hsu, F., Güvener, T., Carl, M.A., Trinidad, R.S., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Rachael, L., Li, M., Schwarz, S., Wang, W.Y., Merz, A.J., David, R., Mougous, J.D., 2010. A Type VI Secretion System of Pseudomonas aeruginosa Targets a Toxin to Bacteria. Cell 7, 25–37. https://doi.org/10.1016/j.chom.2009.12.007.A
Houben, E.N.G., Korotkov, K. V., Bitter, W., 2014. Take five - Type VII secretion systems of Mycobacteria. Biochim. Biophys. Acta - Mol. Cell Res. 1843, 1707–1716. https://doi.org/10.1016/j.bbamcr.2013.11.003
Hsieh, P.F., Lu, Y.R., Lin, T.L., Lai, L.Y., Wang, J.T., 2019. Klebsiella pneumoniae Type VI Secretion System Contributes to Bacterial Competition, Cell Invasion, Type-1 Fimbriae Expression, and in Vivo Colonization. J. Infect. Dis. 219, 637–647. https://doi.org/10.1093/infdis/jiy534
Hutchings, M., Truman, A., Wilkinson, B., 2019. Antibiotics: past, present and future. Curr. Opin. Microbiol. 51, 72–80. https://doi.org/10.1016/j.mib.2019.10.008
Illergård, K., Ardell, D.H., Elofsson, A., 2009. Structure is three to ten times more conserved than sequence - A study of structural response in protein cores. Proteins Struct. Funct. Bioinforma. 77, 499–508. https://doi.org/10.1002/prot.22458
Imai, D.M., Scott Samuels, D., Feng, S., Hodzic, E., Olsen, K., Barthold, S.W., 2013. The early dissemination defect attributed to disruption of decorin-binding proteins is abolished in chronic murine lyme borreliosis. Infect. Immun. 81, 1663–1673. https://doi.org/10.1128/IAI.01359-12
Iwasaki, A., Medzhitov, R., 2015. Control of adaptive immunity by the innate immune system. Nat. Immunol. 16, 343–353. https://doi.org/10.1038/ni.3123
Jewett, M.W., Lawrence, K.A., Bestor, A., Byram, R., Gherardini, F., Rosa, P.A., 2009. GuaA and GuaB are essential for Borrelia burgdorferi survival in the tick-mouse infection cycle. J. Bacteriol. https://doi.org/10.1128/JB.00450-09
Jones, A.M., Low, D.A., Hayes, C.S., 2017. Can't you hear me knocking: contact-dependent competition and cooperation in bacteria. Emerg. Top. Life Sci. 1, 75–83. https://doi.org/10.1042/etls20160019
Jones, D.T., 1999. Protein secondary

structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202. https://doi.org/10.1006/jmbi.1999.3091

Jumper, J., Evans, R., Pritzel, A., Green, T., 2020. High Accuracy Protein Structure Prediction Using Deep Learning [WWW Document]. https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology.

Juncker, A., Willenbrock, H., 2003. Prediction of lipoprotein signal peptides in Gram negative bacteria. Protein Sci. 12, 1652–1662. https://doi.org/10.1110/ps.0303703.Biosynthesis

Jurėnas, D., Journet, L., 2020. Activity, delivery, and diversity of Type VI secretion effectors. Mol. Microbiol. 1–12. https://doi.org/10.1111/mmi.14648

Käll, L., Krogh, A., Sonnhammer, E.L.L., 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res. 35, 429–432. https://doi.org/10.1093/nar/gkm256

Kallmann, B.A., Wagner, S., Hummel, V., Buttmann, M., Bayas, A., Tonn, J.C., Rieckmann, P., 2002. Characteristic gene expression profile of primary human cerebral endothelial cells. FASEB J. 16, 589–591. https://doi.org/10.1096/fj.01-0594fje

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30.

Kapoor, G., Saigal, S., Elongavan, A., 2017. Action and resistance mechanisms of antibiotics: A guide for clinicians. J. Anaesthesiol. Clin. Pharmacol. 33, 300–305.

Kaur, B.P., Secord, E., 2019. Innate Immunity. Pediatr. Clin. North Am. 66, 905–911. https://doi.org/10.1016/j.pcl.2019.06.011

Kawasaki, K., Ernst, R.K., Miller, S.I., 2004. 3-O-Deacylation of Lipid A by PagL, a PhoP/PhoQ-regulated Deacylase of Salmonella typhimurium, Modulates Signaling through Toll-like Receptor 4. J. Biol. Chem. 279, 20044–20048. https://doi.org/10.1074/jbc.M401275200

Kc, D.B., 2017. Recent advances in sequence-based protein structure prediction. Brief. Bioinform. 18, 1021–1032. https://doi.org/10.1093/bib/bbw070

Kerstholt, M., Netea, M.G., Joosten, L.A.B., 2020. Borrelia burgdorferi hijacks cellular metabolism of immune cells: Consequences for host defense. Ticks Tick. Borne. Dis. 101386. https://doi.org/10.1016/j.ttbdis.2020.101386

Kilstrup, M., Hammer, K., Jensen, P.R., Martinussen, J., 2005. Nucleotide metabolism and its control in lactic acid bacteria. FEMS Microbiol. Rev. 29, 555–590. https://doi.org/10.1016/j.femsre.2005.04.006

Kim, J., Thanabalasuriar, A., Chaworth-Musters, T., Fromme, J.C., Frey, E.A., Lario, P.I.I., Metalnikov, P., Rizg, K., Thomas, N.A., Lee, S.F., Hartland, E.L., Hardwidge, P.R.R., Pawson, T., Strynadka, N.C., Finlay, B.B., Schekman, R., Gruenheid, S., 2007. The Bacterial Virulence Factor NleA Inhibits Cellular Protein Secretion by Disrupting Mammalian COPII Function. Cell Host Microbe 2, 160–171. https://doi.org/10.1016/j.chom.2007.07.010

Klein, T.A., Ahmad, S., Whitney, J.C., 2020. Contact-Dependent Interbacterial Antagonism Mediated by Protein Secretion Machines. Trends Microbiol. 28, 387–400. https://doi.org/10.1016/j.tim.2020.01.003

Klemm, P., Schembri, M.A., 2000. Bacterial adhesins: Function and structure. Int. J. Med. Microbiol. 290, 27–35. https://doi.org/10.1016/S1438-4221(00)80102-2

Korotkov, K. V., Sandkvist, M., 2019. Architecture, Function, and Substrates of the Type II Secretion System. Protein Secret. Bact. 227–244. https://doi.org/10.1128/ecosalplus.e

sp-0034-2018
Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., Moult, J., 2019. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins Struct. Funct. Bioinforma. 87, 1011–1020. https://doi.org/10.1002/prot.25823
Kufareva, I., Abagyan, R., 2012. Methods of protein structure comparison. Methods Mol. Biol. 857, 231–257. https://doi.org/10.1007/978-1-61779-588-6_10
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054
Lasica, A.M., Ksiazek, M., Madej, M., Potempa, J., 2017. The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. Front. Cell. Infect. Microbiol. 7. https://doi.org/10.3389/fcimb.2017.00215
Lawlor, M.S., Handley, S.A., Miller, V.L., 2006. Comparison of the host responses to wild-type and cpsB mutant Klebsiella pneumoniae infections. Infect. Immun. 74, 5402–5407. https://doi.org/10.1128/IAI.00244-06
Lawrence, K.A., Jewett, M.W., Rosa, P.A., Gherardini, F.C., 2009. Borrelia burgdorferi bb0426 encodes a 2′-deoxyribosyltransferase that plays a central role in purine salvage. Mol. Microbiol. https://doi.org/10.1111/j.1365-2958.2009.06740.x
Le, S.Q., Gascuel, O., 2008. An Improved General Amino Acid Replacement Matrix. Mol. Biol. Evol. 25, 1307–1320. https://doi.org/10.1093/molbev/msn067
Lehman, M.K., Nuxoll, A.S., Yamada, K.J., Kielian, T., Carson, S.D., Fey, P.D., 2019. Protease-Mediated Growth of Staphylococcus aureus on Host Proteins Is opp3 Dependent 10, 1–17.
Lehtonen, J. V, Still, D.-J., Rantanen, V.-V., Ekholm, J., Björklund, D., Iftikhar, Z., Huhtala, M., Repo, S., Jussila, A., Jaakkola, J., Pentikäinen, O., Nyrönen, T., Salminen, T., Gyllenberg, M., Johnson, M.S., 2004. BODIL: a molecular modeling environment for structure-function analysis and drug design. J. Comput. Aided. Mol. Des. 18, 401–19.
Leiman, P.G., Basler, M., Ramagopal, U. a, Bonanno, J.B., Sauder, J.M., Pukatzki, S., Burley, S.K., Almo, S.C., Mekalanos, J.J., 2009. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proc. Natl. Acad. Sci. U. S. A. 106, 4154–9. https://doi.org/10.1073/pnas.0813360106
Leong, J.M., Wang, H., Magoun, L., Field, J.A., Morrissey, P.E., Robbins, D., Tatro, J.B., Coburn, J., Parveen, N., 1998. Different classes of proteoglycans contribute to the attachment of Borrelia burgdorferi to cultured endothelial and brain cells. Infect. Immun. 66, 994–999. https://doi.org/10.1128/iai.66.3.994-999.1998
Lescot, M., Audic, S., Robert, C., Nguyen, T.T., Blanc, G., Cutler, S.J., Wincker, P., Couloux, A., Claverie, J.M., Raoult, D., Drancourt, M., 2008. The genome of Borrelia recurrentis, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne Borrelia duttonii. PLoS Genet. 4. https://doi.org/10.1371/journal.pgen.1000185
Licht, A., Schneider, E., 2011. ATP binding cassette systems: structures, mechanisms, and functions. Cent. Eur. J. Biol. 6, 785–801. https://doi.org/10.2478/s11535-011-0054-4
Lin, Y.P., Benoit, V., Yang, X., Martínez-Herranz, R., Pal, U., Leong, J.M., 2014. Strain-Specific Variation of the Decorin-Binding Adhesin DbpA Influences the Tissue Tropism of the Lyme Disease Spirochete. PLoS Pathog. 10. https://doi.org/10.1371/journal.ppat.1004238
Lin, Y.P., Li, L., Zhang, F., Linhardt, R.J., 2017. Borrelia burgdorferi

glycosaminoglycan-binding proteins: A potential target for new therapeutics against Lyme disease. Microbiol. (United Kingdom) 163, 1759–1766. https://doi.org/10.1099/mic.0.000571

Lin, Y.P., Tan, X., Caine, J.A., Castellanos, M., Chaconas, G., Coburn, J., Leong, J.M., 2020. Strain-specific joint invasion and colonization by Lyme disease spirochetes is promoted by outer surface protein C. PLoS Pathog. 16, 1–29. https://doi.org/10.1371/journal.ppat.1008516

Llobet, E., Martínez-Moliner, V., Moranta, D., Dahlström, K.M., Regueiro, V., Tomása, A., Cano, V., Pérez-Gutiérrez, C., Frank, C.G., Fernández-Carrasco, H., Insua, J.L., Salminen, T.A., Garmendia, J., Bengoechea, J.A., 2015. Deciphering tissue-induced Klebsiella pneumoniae lipid a structure. Proc. Natl. Acad. Sci. U. S. A. 112, E6369–E6378. https://doi.org/10.1073/pnas.1508820112

Maghrabi, A.H.A., Mcguffin, L.J., 2017. ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. Nucleic Acids Res. 45. https://doi.org/10.1093/nar/gkx332

March, C., Cano, V., Moranta, D., Llobet, E., Pérez-Gutiérrez, C., Tomás, J.M., Suárez, T., Garmendia, J., Bengoechea, J.A., 2013. Role of Bacterial Surface Structures on the Interaction of Klebsiella pneumoniae with Phagocytes. PLoS One 8, 1–16. https://doi.org/10.1371/journal.pone.0056847

Matsuura, M., 2013. Structural modifications of bacterial lipopolysaccharide that facilitate gram-negative bacteria evasion of host innate immunity. Front. Immunol. 4, 1–9. https://doi.org/10.3389/fimmu.2013.00109

Mccoy, L.S., Xie, Y., Tor, Y., 2011. Antibiotics that target protein synthesis. Wiley Interdiscip. Rev. RNA 2, 209–232. https://doi.org/10.1002/wrna.60

McGillivray, S.M., Ebrahimi, C.M., Fisher, N., Sabet, M., Zhang, D.X., Chen, Y., Haste, N.M., Aroian, R. V., Gallo, R.L., Guiney, D.G., Friedlander, A.M., Koehler, T.M., Nizet, V., 2009. ClpX contributes to innate defense peptide resistance and virulence phenotypes of Bacillus anthracis. J. Innate Immun. 1, 494–506. https://doi.org/10.1159/000225955

Merino, S., Camprubi, S., Alberti, S., Benedi, V.J., Tomas, J.M., 1992. Mechanisms of Klebsiella pneumoniae resistance to complement-mediated killing. Infect. Immun. 60, 2529–2535. https://doi.org/10.1128/iai.60.6.2529-2535.1992

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.Y., Finn, R.D., 2019. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 47, D351–D360. https://doi.org/10.1093/nar/gky1100

Mittl, P.R.E., Schneider-Brachert, W., 2007. Sel1-like repeat proteins in signal transduction. Cell. Signal. 19, 20–31. https://doi.org/10.1016/j.cellsig.2006.05.034

Morgan, A., Sepuru, K.M., Feng, W., Rajarathnam, K., Wang, X., 2015. Flexible Linker Modulates Glycosaminoglycan Affinity of Decorin Binding Protein A. Biochemistry 54, 5113–5119. https://doi.org/10.1021/acs.biochem.5b00253

Morgan, A.M., Wang, X., 2015. Structural mechanisms underlying sequence-dependent variations in GAG affinities of decorin binding protein A, a Borrelia burgdorferi adhesin. Biochem. J. 467, 439–451.

https://doi.org/10.1042/BJ20141201
Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., Goodman, A.L., Joachimiak, G., Ordoñez, C.L., Walz, T., Joachimiak, A., Mekalanos, J.J., 2006. A virulence locus of Psedomonas aeruginosa encodes a protein secretion apparatus 312, 1526–1530.
Munita, J.M., Arias, C.A., 2016. Mechanisms of antibiotic resistance. Microb. Spectr. 4, 1–24. https://doi.org/10.3389/fmicb.2015.00034
Murdoch, S.L., Trunk, K., English, G., Fritsch, M.J., Pourkarimi, E., Coulthurst, S.J., 2011. The opportunistic pathogen Serratia marcescens utilizes type VI secretion to target bacterial competitors. J. Bacteriol. 193, 6057–6069. https://doi.org/10.1128/JB.05671-11
Nguyen, G.T., Green, E.R., Mecsas, J., 2017. Neutrophils to the ROScue: Mechanisms of NADPH oxidase activation and bacterial resistance. Front. Cell. Infect. Microbiol. 7. https://doi.org/10.3389/fcimb.2017.00373
Nguyen, L.T., Haney, E.F., Vogel, H.J., 2011. The expanding scope of antimicrobial peptide structures and their modes of action. Trends Biotechnol. 29, 464–472. https://doi.org/10.1016/j.tibtech.2011.05.001
Nicholson, L.B., 2016. The immune system. Essays Biochem. 60, 275–301. https://doi.org/10.1042/EBC20160017
Norris, S.J., 2015. vls Antigenic Variation Systems of Lyme Disease Borrelia : Eluding Host Immunity through both Random, Segmental Gene Conversion and Framework Heterogeneity . Mob. DNA III 471–489. https://doi.org/10.1128/9781555819217.ch22
Núñez, G., Sakamoto, K., Soares, M.P., 2018. Innate Nutritional Immunity. J. Immunol. 201, 11–18. https://doi.org/10.4049/jimmunol.1800325
Paczosa, M.K., Mecsas, J., 2016. Klebsiella pneumoniae: Going on the Offense with a Strong Defense. Microbiol. Mol. Biol. Rev. 80, 629–661. https://doi.org/10.1128/mmbr.00078-15
Palmer, G.H., Bankhead, T., Seifert, H.S., 2016. Antigenic Variation in Bacterial Pathogens. Microbiol spectr 4.
Parsons, N.J., Boons, G.J., Ashton, P.R., Redfern, P.D., Quirk, P., Gao, Y., Constantinidou, C., Patel, J., Bramley, J., Cole, J.A., Smith, H., 1996. Lactic acid is the factor in blood cell extracts which enhances the ability of CMP-NANA to sialylate gonococcal lipopolysaccharide and induce serum resistance. Microb. Pathog. 20, 87–100. https://doi.org/10.1006/mpat.1996.0008
Passalacqua, K.D., Charbonneau, M.-E., O'Riordan, M.X.D., 2016. Bacterial Metabolism Shapes the Host–Pathogen Interface. Microbiol. Spectr. 4, 1–21. https://doi.org/10.1128/microbiolspec.vmbf-0027-2015
Patel, S., Akhtar, N., 2017. Antimicrobial peptides (AMPs): The quintessential 'offense and defense' molecules are more than antimicrobials. Biomed. Pharmacother. 95, 1276–1283. https://doi.org/10.1016/j.biopha.2017.09.042
Patodia, S., 2014. Molecular Dynamics Simulation of Proteins: A Brief Overview. J. Phys. Chem. Biophys. 4, 4–7. https://doi.org/10.4172/2161-0398.1000166
Paton, J.C., Trappetti, C., 2019. Streptococcus pneumoniae Capsular Polysaccharide. Microbiol. Spectr. 7, 1–15. https://doi.org/10.1128/microbiolspec.gpp3-0019-2018
Patro, L.P.P., Rathinavelan, T., 2019. Targeting the Sugary Armor of Klebsiella Species. Front. Cell. Infect. Microbiol. 9. https://doi.org/10.3389/fcimb.2019.00367
Patz, S., Becker, Y., Richert-Pöggeler, K.R., Berger, B., Ruppel, S., Huson, D.H., Becker, M., 2019. Phage tail-like particles are versatile bacterial

nanomachines – A mini-review. J. Adv. Res. 19, 75–84. https://doi.org/10.1016/j.jare.2019.04.003

Peabody, M.A., Laird, M.R., Vlasschaert, C., Lo, R., Brinkman, F.S.L., 2016. PSORTdb: Expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. Nucleic Acids Res. 44, D663–D668. https://doi.org/10.1093/nar/gkv1271

Peraro, M.D., Van Der Goot, F.G., 2016. Pore-forming toxins: Ancient, but never really out of fashion. Nat. Rev. Microbiol. 14, 77–92. https://doi.org/10.1038/nrmicro.2015.3

Perez-Riba, A., Itzhaki, L.S., 2019. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. Curr. Opin. Struct. Biol. 54, 43–49. https://doi.org/10.1016/j.sbi.2018.12.004

Pettersson, J., Schrumpf, M.E., Raffel, S.J., Porcella, S.F., Guyard, C., Lawrence, K., Gherardini, F.C., Schwan, T.G., 2007. Purine salvage pathways among Borrelia species. Infect. Immun. https://doi.org/10.1128/IAI.00199-07

Petzke, M., Schwartz, I., 2015. Borrelia burgdorferi Pathogenesis and the Immune Response. Clin. Lab. Med. 35, 745–764. https://doi.org/10.1016/j.cll.2015.07.004

Pizarro-Cerdá, J., Cossart, P., 2006. Bacterial adhesion and entry into host cells. Cell 124, 715–727. https://doi.org/10.1016/j.cell.2006.02.012

Pohanka, M., 2013. Role of oxidative stress in infectious diseases. A review. Folia Microbiol. (Praha). 58, 503–513. https://doi.org/10.1007/s12223-013-0239-5

Posey, E.J., Gherardini, F.C., 2000. Lack of a Role for Iron in the Lyme Disease Patho 288, 245–261.

Price, C.T.D., Al-quadan, T., Santic, M., Rosenshine, I., 2011. Host Proteasomal Degradation generate amino acids essential for intracellular bacterial growth 334.

Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., Mekalanos, J.J., 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc. Natl. Acad. Sci. U. S. A. 104, 15508–15513. https://doi.org/10.1073/pnas.0706532104

Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg, J.F., Mekalanos, J.J., 2006. Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. Proc. Natl. Acad. Sci. 103, 1528–1533. https://doi.org/10.1073/pnas.0510322103

Radolf, J.D., Caimano, M.J., Stevenson, B., Hu, L.T., 2012. Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. Nat. Rev. Microbiol. 10, 87–99. https://doi.org/10.1038/nrmicro2714

Ramamoorthy, R., McClain, N.A., Gautam, A., Scholl-Meeker, D., 2005. Expression of the bmpB Gene of Borrelia burgdorferi is modulated by two distinct transcription termination events. J. Bacteriol. 187, 2592–2600. https://doi.org/10.1128/JB.187.8.2592-2600.2005

Ramamoorthy, R., Povinelli, L., Philipp, M.T., 1996. Molecular characterization, genomic arrangement, and expression of bmpD, a new member of the bmp class of genes encoding membrane proteins of Borrelia burgdorferi. Infect. Immun. 64, 1259–1264.

Rameshwaram, N.R., Singh, P., Ghosh, S., Mukhopadhyay, S., 2018. Lipid metabolism and intracellular bacterial virulence: Key to next-generation therapeutics. Future Microbiol. 13, 1301–1328. https://doi.org/10.2217/fmb-2018-0013

Records, A.R., 2011. The Type VI Secretion System: A Multipurpose Delivery System with a Phage-Like Machinery. Mol. Plant-Microbe Interact. 24, 751–

757. https://doi.org/10.1094/MPMI-11-10-0262
Reddick, L.E., Alto, N.M., 2014. Bacteria fighting back: How pathogens target and subvert the host innate immune system. Mol. Cell 54, 321–328. https://doi.org/10.1016/j.molcel.2014.03.010
Rees, D.C., Johnson, E., Lewinson, O., 2009. ABC transporters: the power to change. Nat. Rev. Mol. Cell Biol. 10, 218–227. https://doi.org/10.1038/nrm2646
Regueiro, V., Moranta, D., Frank, C.G., Larrarte, E., Margareto, J., March, C., Garmendia, J., Bengoechea, J.A., 2011. Klebsiella pneumoniae subverts the activation of inflammatory responses in a NOD1-dependent manner. Cell. Microbiol. 13, 135–153. https://doi.org/10.1111/j.1462-5822.2010.01526.x
Reniere, M.L., 2018. Reduce, induce, thrive: Bacterial redox sensing during pathogenesis. J. Bacteriol. 200, 1–12. https://doi.org/10.1128/JB.00128-18.
Ribet, D., Cossart, P., 2015. How bacterial pathogens colonize their hosts and invade deeper tissues. Microbes Infect. 17, 173–183. https://doi.org/10.1016/j.micinf.2015.01.004
Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42. https://doi.org/10.1093/nar/gku316
Rohmer, L., Hocquet, D., Miller, S.I., 2011. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. Trends Microbiol. 19, 341–348. https://doi.org/10.1016/j.tim.2011.04.003
Rost, B., Sander, C., 1996. Sequence-Structure Gap By Structure Predictions. Annu. Rev. Biophys. Biomol. Struct. 25, 113–136.
Rostand, K.S., Esko, J.D., 1997. Microbial adherence to and invasion through proteoglycans. Infect. Immun. 65, 1–8. https://doi.org/10.1128/iai.65.1.1-8.1997
Roy, A., Kucukural, A., Zhang, Y., 2011. I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc 5, 725–738. https://doi.org/10.1038/nprot.2010.5.I-TASSER
Ruer, S., Pinotsis, N., Steadman, D., Waksman, G., Remaut, H., 2015. Virulence-targeted Antibacterials: Concept, Promise, and Susceptibility to Resistance Mechanisms. Chem. Biol. Drug Des. 86, 379–399. https://doi.org/10.1111/cbdd.12517
Ruhe, Z.C., Low, D.A., Hayes, C.S., 2020. Polymorphic Toxins and Their Immunity Proteins: Diversity, Evolution, and Mechanisms of Delivery. Annu. Rev. Microbiol. 74, 497–520. https://doi.org/10.1146/annurev-micro-020518-115638
Russell, A.B., Peterson, S.B., Mougous, J.D., 2014. Type VI secretion system effectors: poisons with a purpose. Nat. Rev. Microbiol. 12, 137–48. https://doi.org/10.1038/nrmicro3185
Sadowski, M.I., Jones, D.T., 2009. The sequence-structure relationship and protein function prediction. Curr. Opin. Struct. Biol. 19, 357–362. https://doi.org/10.1016/j.sbi.2009.03.008
Saier Jr., M.H., Paulsen, I.T., 2000. Whole genome analyses of transporters in spirochetes: Borrelia burgdorferi and Treponema pallidum. J Mol Microbiol Biotechnol 2, 393–9.
Šali, A., Blundell, T.L., 1993. Comparative Protein Modelling by Satisfaction of Spatial Restraints. J. Mol. Biol. 234, 779–815. https://doi.org/10.1006/jmbi.1993.1626
Salo, J., Jaatinen, A., Söderström, M., Viljanen, M.K., Hytönen, J., 2015. Decorin binding proteins of Borrelia burgdorferi promote arthritis development and joint specific post-treatment DNA persistence in mice. PLoS One 10, 1–17. https://doi.org/10.1371/journal.pone.0121512
Salo, J., Loimaranta, V., Lahdenne, P., Viljanen, M.K., Hytönen, J., 2011. Decorin binding by DbpA and B of

Borrelia garinii, Borrelia afzelii, and Borrelia burgdorferi sensu stricto. J. Infect. Dis. 204, 65–73. https://doi.org/10.1093/infdis/jir207
Sanada, T., Kim, M., Mimuro, H., Suzuki, M., Ogawa, M., Oyama, A., Ashida, H., Kobayashi, T., Koyama, T., Nagai, S., Shibata, Y., Gohda, J., Inoue, J.I., Mizushima, T., Sasakawa, C., 2012. The Shigella flexneri effector OspI deamidates UBC13 to dampen the inflammatory response. Nature 483, 623–626. https://doi.org/10.1038/nature10894
Scheepers, G.H., Lycklama a Nijeholt, J.A., Poolman, B., 2016. An updated structural classification of substrate-binding proteins. FEBS Lett. 590, 4393–4401. https://doi.org/10.1002/1873-3468.12445
Schnarr, S., Franz, J.K., Krause, A., Zeidler, H., 2006. Lyme borreliosis. Best Pract. Res. Clin. Rheumatol. 20, 1099–1118. https://doi.org/10.1016/j.berh.2006.08.006
Schwarz, S., West, T.E., Boyer, F., Chiang, W.C., Carl, M.A., Hood, R.D., Rohmer, L., Tolker-Nielsen, T., Skerrett, S.J., Mougous, J.D., 2010. Burkholderia type vi secretion systems have distinct roles in eukaryotic and bacterial cell interactions. PLoS Pathog. 6, 77–78. https://doi.org/10.1371/journal.ppat.1001068
Sen, R., Nayak, L., De, R.K., 2016. A review on host–pathogen interactions: classification and prediction. Eur. J. Clin. Microbiol. Infect. Dis. 35, 1581–1599. https://doi.org/10.1007/s10096-016-2716-7
Sgro, G.G., Oka, G.U., Souza, D.P., Cenens, W., Bayer-Santos, E., Matsuyama, B.Y., Bueno, N.F., Dos Santos, T.R., Alvarez-Martinez, C.E., Salinas, R.K., Farah, C.S., 2019. Bacteria-killing type IV secretion systems. Front. Microbiol. 10, 1–20. https://doi.org/10.3389/fmicb.2019.01078
Shafer, W.M., Qu, X.D., Waring, A.J., Lehrer, R.I., 1998. Modulation of Neisseria gonorrhoeae susceptibility to vertebrate antibacterial peptides due to a member of the resistance/nodulation/division efflux pump family. Proc. Natl. Acad. Sci. U. S. A. 95, 1829–1833. https://doi.org/10.1073/pnas.95.4.1829
Si, M., Zhao, C., Burkinshaw, B., Zhang, B., Wei, D., Wang, Y., Dong, T.G., Shen, X., 2017. Manganese scavenging and oxidative stress response mediated by type VI secretion system in Burkholderia thailandensis. Proc. Natl. Acad. Sci. 2233–2242. https://doi.org/10.1073/pnas.1614902114
Simpson, W.J., Cieplak, W., Schrumpf, M.E., Barbour, A.G., Schwan, T.G., 1994. Nucleotide sequence and analysis of in Borrelia burgdorferi encoding the P39 antigen the gene immunogenic. FEMS Microbiol. Lett. 119, 381–388. https://doi.org/10.1111/j.1574-6968.1994.tb06917.x
Simpson, W.J., Schrumpf, M.E., Schwan, T.G., 1990. Reactivity of human Lyme borreliosis sera with a 39-kilodalton antigen specific to Borrelia burgdorferi. J. Clin. Microbiol. 28, 1329–37.
Skaar, E.P., 2010. The battle for iron between bacterial pathogens and their vertebrate hosts. PLoS Pathog. 6, 1–2. https://doi.org/10.1371/journal.ppat.1000949
Soltis, D.E., Soltis, P.S., 2014. Update on Molecular Systematics The Role of Phylogenetics in Comparative Genetics 1 132, 1790–1800. https://doi.org/10.1104/pp.103.022509.and
Soltis, D.E., Soltis, P.S., 2003. Applying the Bootstrap in Phylogeny Reconstruction. Stat. Sci. 18, 256–267. https://doi.org/10.1214/ss/1063994980
Song, L., Pan, J., Yang, Y., Zhang, Z., Cui, R., Jia, S., Wang, Z., Yang, C., Xu, L., Dong, T.G., Wang, Y., Shen, X., 2021. Contact-independent killing mediated by a T6SS effector with intrinsic cell-entry properties. Nat. Commun. 12. https://doi.org/10.1038/s41467-

020-20726-8
Sorbara, M.T., Philpott, D.J., 2011. Peptidoglycan: A critical activator of the mammalian immune system during infection and homeostasis. Immunol. Rev. 243, 40–60. https://doi.org/10.1111/j.1600-065X.2011.01047.x
Sperandio, B., Fischer, N., Sansonetti, P.J., 2015. Mucosal physical and chemical innate barriers: Lessons from microbial evasion strategies. Semin. Immunol. 27, 111–118. https://doi.org/10.1016/j.smim.2015.03.011
Spínola-Amilibia, M., Davó-Siguero, I., Ruiz, F.M., Santillana, E., Medrano, F.J., Romero, A., 2016. The structure of VgrG1 from Pseudomonas aeruginosa, the needle tip of the bacterial type VI secretion system. Acta Crystallogr. Sect. D, Struct. Biol. https://doi.org/10.1107/S2059798315021142
Spitz, O., Erenburg, I.N., Beer, T., Kanonenberg, K., Holland, I.B., Schmitt, L., 2019. Type I Secretion Systems—One Mechanism for All? Protein Secret. Bact. 215–225. https://doi.org/10.1128/microbiolspec.psib-0003-2018
Stanek, G., Strle, F., 2018. Lyme borreliosis-from tick bite to diagnosis and treatment. FEMS Microbiol. Rev. 42, 233–258. https://doi.org/10.1093/femsre/fux047
Stones, D.H., Krachler, A.M., 2016. Against the tide: The role of bacterial Adhesion in host colonization. Biochem. Soc. Trans. 44, 1571–1580. https://doi.org/10.1042/BST20160186
Storey, D., McNally, A., Åstrand, M., Santos, J.P.G., Rodriguez-Escudero, I., Elmore, B., Palacios, L., Marshall, H., Hobley, L., Molina, M., Cid, V.J., Salminen, T.A., Bengoechea, J.A., 2020. Klebsiella pneumoniae type VI secretion system-mediated microbial competition is PhoPQ controlled and reactive oxygen species dependent, PLoS Pathogens. https://doi.org/10.1371/journal.ppat.1007969
Stubbendieck, R.M., Straight, P.D., 2016. Multifaceted interfaces of bacterial competition. J. Bacteriol. 198, 2145–2155. https://doi.org/10.1128/JB.00275-16
Talagrand-Reboul, E., Boyer, P.H., Bergström, S., Vial, L., Boulanger, N., 2018. Relapsing fevers: Neglected tick-borne diseases. Front. Cell. Infect. Microbiol. 8. https://doi.org/10.3389/fcimb.2018.00098
Tanaka, K.J., Song, S., Mason, K., Pinkett, H.W., 2018. Selective substrate uptake: The role of ATP-binding cassette (ABC) importers in pathogenesis. Biochim. Biophys. Acta - Biomembr. 1860, 868–877. https://doi.org/10.1016/j.bbamem.2017.08.011
Thakur, A., Mikkelsen, H., Jungersen, G., 2019. Intracellular pathogens: Host immunity and microbial persistence strategies. J. Immunol. Res. 2019. https://doi.org/10.1155/2019/1356540
Ting Liang, F., Kenneth Nelson, F., Fikrig, E., 2002. Brief Definitive Report Molecular Adaptation of Borrelia burgdorferi in the Murine Host. J. Exp. Med 00, 275–280. https://doi.org/10.1084/jem.20020770
Ting, S.Y., Bosch, D.E., Mangiameli, S.M., Radey, M.C., Huang, S., Park, Y.J., Kelly, K.A., Filip, S.K., Goo, Y.A., Eng, J.K., Allaire, M., Veesler, D., Wiggins, P.A., Peterson, S.B., Mougous, J.D., 2018. Bifunctional Immunity Proteins Protect Bacteria against FtsZ-Targeting ADP-Ribosylating Toxins. Cell 175, 1380-1392.e14. https://doi.org/10.1016/j.cell.2018.09.037
Tsirigotaki, A., De Geyter, J., Šoštarić, N., Economou, A., Karamanou, S., 2017. Protein export through the bacterial Sec pathway. Nat. Rev. Microbiol. 15, 21–36. https://doi.org/10.1038/nrmicro.2016.161
Uribe-Querol, E., Rosales, C., 2017. Control of phagocytosis by microbial pathogens. Front. Immunol. 8, 1–23. https://doi.org/10.3389/fimmu.2017.01368

Varki, A., Cummings, R., Esko, J., 1999. Chapter 11 Proteoglycans and Glycosaminoglycans, in: Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, pp. 1–23.
Vascon, F., Gasparotto, M., Giacomello, M., Cendron, L., Bergantino, E., Filippini, F., Righetto, I., 2020. Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design. Comput. Struct. Biotechnol. J. 18, 1774–1789. https://doi.org/10.1016/j.csbj.2020.06.029
Verma, A., Brissette, C.A., Bowman, A., Stevenson, B., 2009. Borrelia burgdorferi BmpA is a laminin-binding protein. Infect. Immun. 77, 4940–4946. https://doi.org/10.1128/IAI.01420-08
Wallner, B., Elofsson, A., 2003. Can correct protein models be identified? Protein Sci. 12, 1073–86. https://doi.org/10.1110/ps.0236803
Wang, T., Si, M., Song, Y., Zhu, W., Gao, F., Wang, Y., Zhang, L., Zhang, W., Wei, G., Luo, Z.Q., Shen, X., 2015. Type VI Secretion System Transports Zn2+ to Combat Multiple Stresses and Host Immunity. PLoS Pathog. 11, 1–25. https://doi.org/10.1371/journal.ppat.1005020
Wang, X., 2012. Solution structure of decorin-binding protein A from Borrelia burgdorferi. Biochemistry 51, 8353–8362. https://doi.org/10.1021/bi3007093
Wang, Y.D., Zhao, S., Hill, C.W., 1998. Rhs elements comprise three subfamilies which diverged prior to acquisition by Escherichia coli. J. Bacteriol. 180, 4102–4110. https://doi.org/10.1128/jb.180.16.4102-4110.1998
Warner, D.F., Evans, J.C., Mizrahi, V., 2014. Nucleotide Metabolism and DNA Replication. Microbiol. Spectr. 2, 1–20. https://doi.org/10.1128/microbiolspec.mgm2-0001-2013
Webb, S.A.R., Kahler, C.M., 2008. Bench-to-bedside review: Bacterial virulence and subversion of host defences. Crit. Care 12, 234. https://doi.org/10.1186/cc7091
Wen, H., Geng, Z., Gao, Z., She, Z., Dong, Y., 2020. Characterization of the Pseudomonas aeruginosa T6SS PldB immunity proteins PA5086, PA5087 and PA5088 explains a novel stockpiling mechanism. Acta Crystallogr. Sect. F Struct. Biol. Commun. 76, 222–227. https://doi.org/10.1107/S2053230X2000566X
Westhoff, S., van Wezel, G.P., Rozen, D.E., 2017. Distance-dependent danger responses in bacteria. Curr. Opin. Microbiol. 36, 95–101. https://doi.org/10.1016/j.mib.2017.02.002
Wettstadt, S., Wood, T.E., Fecht, S., Filloux, A., 2019. Delivery of the Pseudomonas aeruginosa phospholipase effectors PldA and PldB in a VgrG- And H2-T6SS-dependent manner. Front. Microbiol. 10, 1–6. https://doi.org/10.3389/fmicb.2019.01718
WHO, 2017. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis.
Wiederstein, M., Sippl, M.J., 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, 407–410. https://doi.org/10.1093/nar/gkm290
Willett, J.L.E., Gucinski, G.C., Fatherree, J.P., Low, D.A., Hayes, C.S., 2015. Contact-dependent growth inhibition toxins exploit multiple independent cell-entry pathways. Proc. Natl. Acad. Sci. U. S. A. 112, 11341–11346. https://doi.org/10.1073/pnas.1512124112
Wilson, J.W., Schurr, M.J., LeBlanc, C.L., Ramamurthy, R., Buchanan, K.L., Nickerson, C.A., 2002. Mechanisms of bacterial pathogenicity. Postgrad. Med. J. 78, 216–224. https://doi.org/10.1136/pmj.78.918.216
Wood, T.E., Howard, S.A., Förster, A., Nolan, L.M., Manoli, E., Bullen, N.P., Yau,

H.C.L., Hachani, A., Hayward, R.D., Whitney, J.C., Vollmer, W., Freemont, P.S., Filloux, A., 2019a. The Pseudomonas aeruginosa T6SS Delivers a Periplasmic Toxin that Disrupts Bacterial Cell Morphology. Cell Rep. 29, 187-201.e7. https://doi.org/10.1016/j.celrep.2019.08.094

Wood, T.E., Howard, S.A., Wettstadt, S., Filloux, A., 2019b. PAAR proteins act as the 'sorting hat' of the type VI vsecretion system. Microbiol. (United Kingdom) 165, 1208–1218. https://doi.org/10.1099/mic.0.000842

Xu, J., Jiao, F., Yu, L., 2007. Protein structure prediction using threading. Methods Mol. Biol. 413, 91–121. https://doi.org/10.1385/1-59745-574-1:91

Yang, J., Yan, R., Roy, A., Xu, D., J, P., Zhang, Y., 2015. The I-TASSER Suite: Protein structure and function prediction. Nat. Methods 12, 7–8. https://doi.org/10.1038/nmeth.3213

Yang, X., Li, Z., She, Z., Geng, Z., Xu, J., Gao, Z., 2016. Structural analysis of Pseudomonas aeruginosa H3-T6SS immunity proteins 590, 2787–2796. https://doi.org/10.1002/1873-3468.12291

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: Principles and practice. Nat. Rev. Genet. 13, 303–314. https://doi.org/10.1038/nrg3186

Yu, C.-S., Lin, C.-J., Hwang, J.-K., 2004. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n -peptide compositions . Protein Sci. 13, 1402–1406. https://doi.org/10.1110/ps.03479604

Zeytuni, N., Ozyamak, E., Ben-Harush, K., Davidov, G., Levin, M., Gat, Y., Moyal, T., Brik, A., Komeili, A., Zarivach, R., 2011. Self-recognition mechanism of mamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. Proc. Natl. Acad. Sci. U. S. A. 108. https://doi.org/10.1073/pnas.1103367108

Zückert, W.R., 2014. Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond. Biochim. Biophys. Acta - Mol. Cell Res. 1843, 1509–1516. https://doi.org/10.1016/j.bbamcr.2014.04.022

ISBN 978-952-12-4076-8