Western Norway University of Applied Sciences

BACHELOR'S THESIS

Determination of the interaction site between microbial Lrr-proteins and gp340 receptor.

A literature review

Bestemmelse av interaksjonsstedet mellom mikrobielle Lrr-proteiner og gp340 reseptor.

En litteraturgjennomgang

Kareem Amer

Faculty of Engineering and Science

Department of Safety, Chemistry and Biomedical laboratory sciences

Supervisor: Gry Sjøholt

29.05.2020

I confirm that the work is self-prepared and that references/source references to all sources used in the work are provided, cf. Regulation relating to academic studies and examinations at the Western Norway University of Applied Sciences (HVL), § 12-1.

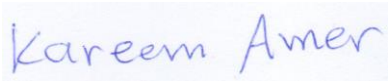
Preface

This bachelor thesis was carried out in the spring of 2020 as a final part of the bachelor's degree program in Biomedical laboratory sciences at Western Norway University of Applied Sciences (HVL) – Bergen campus, in cooperation with Turku University of Applied sciences (TUAS) in Finland. The original project was supposed to be carried out in Turku Bioscience Centre under the supervision of professor Tassos Papageorgiou, but due to COVID-19 pandemic and the consequences that followed, the plan has been changed.

Since it was not possible for me to use the laboratory and other facilities to perform laboratory experiments and analysis, I was given the alternative of writing a literature review about the same topic instead. This was a bit of challenging task, since I was not trained during this educational program to write those types of papers, and of course because I was working on this all alone under stressful circumstances, but I managed to finish this project within the deadline.

In the end, I would like to extend my sincere gratitude and appreciation to those who helped me during this period, especially the supervisor at HVL, Gry Sjøholt and the supervisor in Turku, Tassos Papageorgiou. I am very grateful for everything that I learned from them and for the great guidance they provided while writing my thesis.

Bergen, May 2020



Kareem Amer

Abstract

This literature review focus on recent research studies unveiling the structural interaction between microbial leucine-rich repeats (Lrr) domains and their corresponding protein receptor, gp340 in the host cells. The structural studies can help us to better understand how bacterial infections happen in a way that would make us see the microbial infection problem from another perspective, which can possibly be used to develop an alternative medication that can replace the traditional antibiotics in the future.

X-ray crystallography is an important method used for structural determination of macromolecules including proteins at the atomic level in a way that can help us to understand how those proteins function and interact with one another.

Both the Lrr-domain of the Spy0843 protein in *Streptococcus pyogenes* and the Lrr-binding SRCR domain of human gp340 have been crystallized. Whereas only the atomic structure of the SRCR domain of gp340 has been revealed by X-ray structural studies so far. The detailed atomic structure of bacterial Lrr-domains specific binding to gp340 remains to be characterized.

Interestingly, several studies suggested that some factors can have an inhibitory affect on the binding between Lrr-protein domains and gp340 receptors. Among those factors N-Acetylneuraminic acid and SRCR1A peptide. Building on those findings, similar factors can be used to develop an anti-adhesion treatment that can replace antibiotics in the future.

Keywords: X-ray crystallography, Spy0843, Leucine-rich repeats (Lrr), Scavenger Receptor Cysteine Rich (SRCR), gp340, anti-adhesion treatment

Abstrakt (På Norsk)

Denne litteraturgjennomgangen setter søkelys på nyere forskningsstudier som avduker det strukturelle samspillet mellom mikrobielle leucinrike gjentakelsesdomener (Lrr) og deres tilsvarende proteinreseptor, gp340 i vertscellene. De strukturelle studiene kan gi en bedre forståelse av hvordan bakterieinfeksjoner foregår slik at vi ser på mikrobiell infeksjons problemet fra et annet synsvinkel, som muligens kan benyttes til å utvikle et alternativt behandling som kan erstatte de tradisjonelle antibiotikaene i fremtiden.

Røntgenkrystallografi er en viktig metode som brukes for strukturell bestemmelse av makromolekyler, blant annet proteiner, på atomnivå slik at vi forstår bedre hvordan disse proteinene fungerer og interagerer med hverandre.

Både Lrr-domenet til Spy0843-proteinet i *Streptococcus pyogenes* og det Lrr-bindende SRCR-domene til humant gp340 har blitt krystallisert. Mens kun atomstrukturen i SRCR-domenet til gp340 har blitt avslørt av røntgenstrukturelle studier hittil. Den detaljerte atomstrukturen for bakterielle Lrr-domener spesifikk binding til gp340 har ikke karakterisert ennå.

Interessant nok antydet flere studier at noen faktorer kan ha en hemmende innvirkning på bindingen mellom Lrr-proteindomene og gp340-reseptorer. Blant disse faktorene N-Acetylneuraminic syre og SRCR1A peptid. På bakgrunn av disse funnene kan lignende faktorer benyttes til å utvikle en antiheftet behandling som kan erstatte antibiotika i fremtiden.

Stikkord: røntgenkrystallografi, Spy0843, leucinrike gjentakelser (Lrr), Scavenger Receptor Cysteine Rich (SRCR), gp340, antiheftet behandling

Table of contents

Preface 1

Abstract 2

Abstrakt (På Norsk) 3

1. Introduction 6

1.1. Methods for structural determination 6

1.2. The basic principle of X-ray crystallography 7

1.3. Lrr protein motifs and their interaction with binding receptors as sites for bacterial infection 7

1.4. Problem statement 9

1.5. Study purpose 9

2. Methods 10

2.1. Literature search 10

3. Results 11

3.1. Chosen articles 11

3.2. Review of the first article 12

3.2.1. The purpose of the research 12

3.2.2. Method 12

3.2.3. Main findings 13

3.3. Review of the second article 14

3.3.1. The purpose of the research 15

3.3.2. Method 15

3.3.3. Main findings 15

3.4. Review of the third article 17

3.4.1. The purpose of the research 17

3.4.2. Method 17

3.4.3. Main findings 18

3.5. Review of the fourth article 20

3.5.1. The purpose of the research 20

3.5.2. Method 20

3.5.3. Main findings 21

3.6. Review of the fifth article 22

3.6.1. The purpose of the research 22
3.6.2. Method 22
3.6.3. Main findings 23
4. Discussion 25
5. References 28

1. Introduction

The biological macromolecules make up most of the dry mass in the living cells. These molecules are consisting of smaller subunits linked together by bonds that formed during an enzyme-catalyzed reaction. There are different types of macromolecules such as DNA and RNA that consists of several nucleotides, polysaccharides that consists of monosaccharides and proteins that consists of amino acids. (Alberts et al., 2014, p.113-114)

1.1. Methods for structural determination

The structure determination of macromolecules is an important analysis that is required in order to have a full understanding of the biological processes that takes place inside the cells. Proteins are the largest group of biological macromolecules and they take part in almost all chemical reactions inside the living cells and in the interactions between cells. It is therefore vital to determine the three-dimensional structure of proteins in order to get a better understanding of how they function. (Papageorgiou & Mattsson, 2014)

There are three main techniques that can be used for structural determination of macromolecules. The first is nuclear magnetic resonance (NMR) spectroscopy. The data collected from NMR method can be recorded in solutions, that can be useful for the structural determination of different proteins and nucleic acids while they perform their functions in the body fluid (Wüthrich, 2003). The problem with this technique is that it can only be used with proteins that are smaller than 25 kDa, and therefore we can't take it in use when we try to determine the structure of complex proteins with large molecular weights. (Frueh et al., 2013).

The second main technique used for determination of the three-dimensional structure of proteins is cryo-electron microscopy (Cryo-EM). Similar with NMR, this technique doesn't require crystallization of proteins. Cryo-EM can give us a chance to look at proteins while they interact with one another and perform their functions (Wang & Wang, 2017).

The third main technique that can be used for three-dimensional structure determination of proteins is X-ray crystallography. This technique can help us to determine protein structure at the atomic level and by that we can collect lots of information that helps to understand the function of proteins and their life processes. This can give us some ideas of how to modify

proteins in order to design new inhibitors as potential drugs and also to initiate more efficient and stable enzymes. (Papageorgiou & Mattsson, 2014).

1.2. The basic principle of X-ray crystallography

As mentioned earlier, X-ray crystallography is a technique that can be used for three-dimensional structure determination of macromolecules at the atomic resolution. Determination of the protein structure using X-ray crystallography requires purification of the protein to more than 95 % purity in sufficient quantities that is enough for setting up crystallizations.

The principle for this technique is based on the fact that X-rays get diffracted by crystals. When a crystal gets exposed to an X-ray beam that will cause the X-ray to diffract into different directions. Based on the pattern of diffraction that acquired from the X-ray scattering from the periodic composition of atoms in the crystal, the electron density can be constructed. The data collected from this process can be used to build an accurate molecular structure model (Papageorgiou & Mattsson, 2014).

1.3. Lrr protein motifs and their interaction with binding receptors as sites for bacterial infection

Leucine-rich repeats (Lrr) are short sequence motifs found in proteins with different cellular functions and locations. Proteins that contains Lrr-motifs are involved in protein-protein interactions. It has been revealed that the crystal structure of ribonuclease inhibitor protein contains leucine-rich repeats that are correspond to β-α structural units. These units are arranged in a way that they form parallel β-sheets with one surface exposed to the solvent, so that protein acquires nanoglobular shape. Such characteristics are thought to be responsible for the binding functions that exist in proteins which contains Lrr motifs. (Kobe & Deisenhofer, 1994)

This literature study will focus at Lrr domains that exist in Spy0843 protein which is found in the gram-positive *Streptococcus pyogenes* (illustrated in Figure 1 a) and other proteins that exist in other pathogenic bacterial species (illustrated in Figure 1 b).

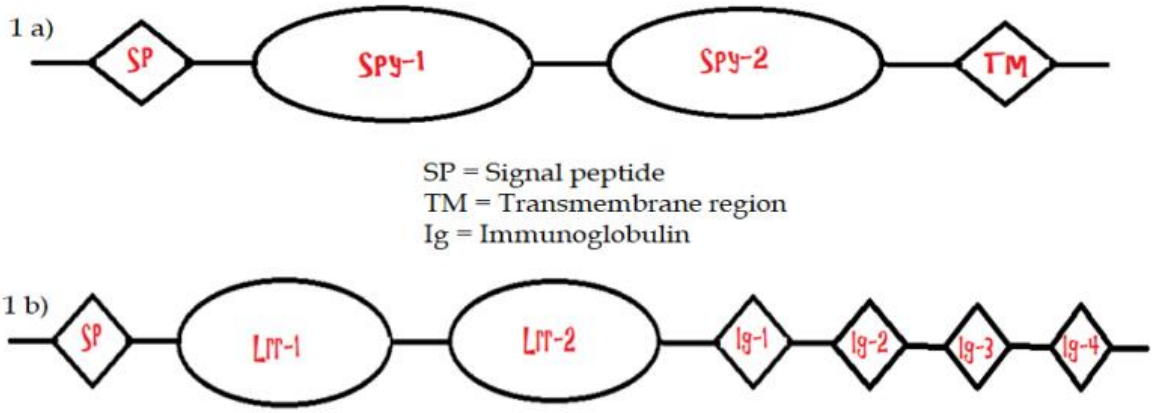


Figure 1 a) Spy0843 from Streptococcus pyogenes b) the Lrr protein domains from Tannerella forsythia. (Tassos Papageorgiou)

The microbial Lrr protein motif 's interacts with the gp340 receptor which is a member of the scavenger receptor cysteine rich (SRCR) superfamily and is found in secretions and at cell surfaces. The Lrr motif and gp340 interaction is considered a key step for bacterial colonization and formation of biofilms. Gp340 is expressed by epithelial cells and cells of the immune system, where its expression is up-regulated after inflammatory stimuli. Studying the structure of microbial Lrr domains can help us to better understand how bacterial infections happen.

Studying such interactions between the microbial Lrr and gp340 might also have some special importance nowadays due to the increase in the bacterial resistance to antibiotics. Structural studies of these interactions can help us to look at the microbial infection problems from another perspective that could lead us to create a solution based on anti-adhesion compound to prevent the binding between Lrr protein domains and gp340 receptors and, thus, the microbial infections (Loimaranta et al., 2009).

In other words, studying such interactions can help us to create an alternative medicine that can replace antibiotics in the future. This medicine would prevent the microbial infections by making it impossible for bacterial cells to bind to their target cells in human body, and by that bacteria will not be able to reproduce and colonize in the host.

1.4. Problem statement

The discovery of Penicillin in 1928 changed the course of medicine and helped us to cure severe illnesses that was considered fatal before such as pneumonia, tuberculosis, meningitis, peritonitis and syphilis. (Hare, 1982, p.1-24)

Since then the antibiotics have been used to cure different forms of bacterial infection, but later it has been discovered that some bacterial species develop a mechanism that make them able to establish an infection despite of antibiotic treatment. This mechanism is referred to as antibiotic resistance and was first discovered in 1940s in Penicillin resistant *Staphylococcus aureus.* This problem has been increased by time as it was discovered different bacterial species that are resistant to antibiotics and that might put our lives again at risk of the illnesses that we thought we manage to get rid of it. (Lobanovska & Pilla, 2017)

In the light of that, we need to find an alternative for antibiotics that can help us in the future when we are about to lose this battle completely to bacteria. By studying the interaction between the microbial Lrr-protein and gp340-receptor we can look at the microbial infection problems from another perspective and that could lead us to come up with a solution based on anti-adhesion compound that can stop the microbial infections by blocking the binding site in gp340 receptor.

1.5. Study purpose

The purpose of this literature study is to provide foundation of knowledge on protein structure determination of the interaction between bacterial Lrr protein domains and their human receptors. The literature study also aims at exploring the different scientific publications in this area of research with special focus on relevance of anti-adhesion treatment as possible alternative for antibiotics.

2. Methods

The initial plan for this project was to express recombinant Lrr-protein in *Escherichia coli* bacterial cells, purifying, and crystallizing it in order to study its structure and function. Due to COVID-19 pandemic, and the shutdown that followed, it was not possible for me to continue with this plan. I had a discussion with both internal and external supervisors in Finland and Norway, and we decided to turn the bachelor project into a literature study.

2.1. Literature search

The literature search was performed by searching online scientific databases for articles that deals with protein structure determination, X-ray crystallography, bacterial Lrr protein domains, gp340 receptors and the interaction between those two. The relevant articles were chosen and analysed in a systematic manner as a literature review.

In this project I used some online databases to look for articles that explain the defined area of research. First, I used the literature databases that exist within National Center for Biotechnology Information (NCBI) to look for articles that include certain keywords. The keywords which I was looking for were as following: X-ray crystallography + Lrr + gp340. This keyword combination gave 0 articles. Then I changed the keyword combination to the following: X-ray + Lrr + gp340 and that resulted only in 1 article under PubMed Central literature database, which is not sufficient. But this article (Haikarainen et al., 2013) was relevant to the topic, that's why I decided to include it in the literature review.

When I used the keyword combination X-ray crystallography + Lrr excluding gp340, I got 96 articles within PubMed literature database and 466 articles within PubMed Central. After a brief evaluation I found that most of those articles are irrelevant to the project purpose.

It was hard to limit the search results by using more keywords in order to get just those articles that are relevant to the search area. Increasing the keywords gave almost 0 relevant articles, and decreasing it gave huge number of articles, most of it were irrelevant. That is why I decided to adopt another strategy.

First, I decided to include an article which I received from my supervisor in Finland, professor Anastassios Papageorgiou (Loimaranta et al., 2009), who is working in the field of X-ray crystallography with a special interest on Lrr and gp340 interaction.

Then I searched google search engine for the keywords: X-ray crystallography + Lrr + gp340. There were big number of articles that came out of this search. After some readings and evaluation, I decided to pick four articles that I found relevant to the study purpose.

The articles will be reviewed in an ascending order, in which I will start with the article which was published first and finish with the latest.

3. Results

In this part of my thesis I will review the results of the literature search. First, I will present the articles which I found relevant for the project purpose, then, I will review each article separately. The review of each article will be divided into three sections: The purpose of the research, Method and Main findings.

3.1. Chosen articles

The articles which I chose to work on in this literature review are the following:

1. Ikegami et al., 2004:
   Multiple Functions of the Leucine-Rich Repeat Protein LrrA of *Treponema denticola*
2. Loimaranta et al., 2009:
   Leucine-rich Repeats of Bacterial Surface Proteins Serve as Common Pattern Recognition Motifs of Human Scavenger Receptor gp340
3. Kukita et al., 2013:
   Staphylococcus aureus SasA Is Responsible for Binding to the Salivary Agglutinin gp340, Derived from Human Saliva
4. Haikarainen et al., 2013:
   Expression, purification and crystallization of the C-terminal LRR domain of *Streptococcus pyogenes* protein 0843

5. Reichhardt et al., 2020:
   Structures of SALSA/DMBT1 SRCR domains reveal the conserved ligand-binding mechanism of the ancient SRCR fold

3.2. Review of the first article

The title of the first article is "Multiple Functions of the Leucine-Rich Repeat Protein LrrA of *Treponema denticola*". It was written by Akihiko Ikegami, Kiyonobu Honma, Ashu Sharma and Howard K. Kuramitsu and was published in American Society for Microbiology Journal "Infection and Immunity", August 2004.

3.2.1. The purpose of the research

The authors of this study worked with the oral spirochete *Treponema denticola* which is associated with periodontal diseases. This bacterium involves in invasion of oral tissues, production of tissue destructive enzymes, formation of cytotoxic products and immunosuppression of host cell functions.

Previously, they had identified the LrrA-gene which encoding a leucine-rich repeats protein LrrA in ATCC 35405 strain of *Treponema denticola*, but it wasn't identified in the strain ATCC 33520 of the bacterium. In this study they wanted to analyse the functions of LrrA protein in *Treponema denticola*. For this purpose, they constructed a LrrA-inactivated mutant of ATCC 35405 strain and a LrrA gene expression transformant of strain ATCC 33520. Examining these constructs would help to understand the virulence of this type of bacteria.

3.2.2. Method

The researchers started by cultivating ATCC 35405 and 33520 strains of *Treponema denticola*, ATCC 43037 strain of *Tannerella forsythensis*, BspA-defective mutant of *Tannerella forsythensis*, 381 strain of *Porphyromonas gingivalis* and NCTC 11326 strain of *Fusobacterium nucleatum* under anaerobic conditions. The JM109 strain of *Escherichia coli*

was used as a host strain for plasmid construction and expression of fusion proteins. That followed by cloning and sequencing of the LrrA-gene of *Treponema denticola.*

A *Treponema denticola* LrrA mutant was constructed using LrrA gene-defective mutant of the ATCC 35405 strain. Chromosomal DNA of *Treponema denticola* was analysed by southern blot technique. Followed by the expression of LrrA-gene in *Treponema denticola* and overexpression and purification of LrrA protein.

Then they separated the recombinant protein His-LrrAΔN using SDS–12 % PAGE, and visualized it by staining with Coomassie brilliant blur R. They used a gel strip containing a band of His-LrrAΔN protein as an antigen. They injected the protein into chickens to produce antibodies against His-LrrAΔN protein. That followed by performing a western plot analysis of *Treponema denticola* with anti-LrrA antiserum.

The next step was performing an immunogold labelling which began by centrifuging a 3-day cultured *Treponema denticola* and washing it with PBS solution. After some steps (see the article) they observed it using transmission electron microscope.

They also added a bacterial suspension of *Treponema denticola* to HEp-2 epithelial tissue layer which was prepared in the laboratory. They incubated them for an hour at 37 °C before washing the tissue epithelial layer with PBST solution. Then they detected and calculated the attached bacterial cells. Then they detected the coaggregation activity between *Treponema denticola* and heterologous bacteria.

That was followed with binding assays of the LrrA derivatives to the recombinant Bsp70 protein, swarming assay and tissue penetration assay.

In the end they assigned the accession number AY545217 for the nucleotide sequence of LrrA-gene after adding it to the GeneBank database.

3.2.3. Main findings

As a result of cloning and sequencing of the lrrA gene of *Treponema denticola*, the Lrr-sequence was identified in the genome database of *Treponema denticola* ATCC 35405 from TIGR with BspA sequence. The full length Lrr-gen which consists of 1068 base pairs was given the name LrrA. The LrrA protein was suggested to be a lipoprotein with a predicted molecular mass of 32829 Da.

They also found a high similarity between the amino acid sequence of Lrr motifs of LrrA protein and other Lrr proteins in bacterial species. For example, they found 57 % identity with Lrr sequence in BspA protein of *Tannerella forsythensis* and 35 % identity with Lrr sequence in TpLRR of *Treponema pallidum*.

As a result of expression of LrrA protein in *Treponema denticola*, the protein was found be expressed in both ATCC 35405 and ATCC 33520&pKMlrrA strains, but it was defective in the last strain. They also found that the LrrA is a lipoprotein associated with extra cytoplasmic cell fraction of ATCC 35405 strain of *Treponema denticola*.

The attachment of *Treponema denticola* to Hep-2 cell tissue layers suggests that the LrrA protein plays a role in the attachment of this bacterium to epithelial cells. The LrrA protein also found to play a role in coaggregation between *Treponema denticola* and other bacterial species that have Lrr-proteins. This coaggregation is believed to contribute to multispecies biofilms.

The binding of LrrA derivatives from *Treponema denticola* to recombinant BspA from *Tannerella forsythensis* suggested that binding of those proteins is mediated by the N-terminal region of LrrA, and not by interaction of Lrr-domains.

The results also suggest the bacterial ability to penetrate the epithelial cell tissues is correlated with the swarming activity of *Treponema denticola*. It also suggests that the swarming activity can be influenced by the presence or absence of LrrA protein.

3.3. Review of the second article

The title of the second article is “Leucine-rich Repeats of Bacterial Surface Proteins Serve as Common Pattern Recognition Motifs of Human Scavenger Receptor gp340”. It was written by Vuokko Loimaranta, Jukka Hytönen, Arto T. Pulliainen, Ashu Sharma, Jorma Tenovuo, Nicklas Strömberg, and Jukka Finne and was published in Journal of Biological Chemistry, May 2009.

3.3.1. The purpose of the research

The purpose of this study is to shed light on the ligand recognition capabilities of gp340 receptor, by using human pathogen *Streptococcus pyogenes* as model bacterium to identify novel bacterial proteins binding to gp340 receptors. The novel *Streptococcus pyogenes* host interaction is mediated by the bacterial surface protein Spy0843 and the gp340 receptor of the host cells. It's known that gp340 receptor binds to conserved Lrr motifs which is found in Spy0843 and that gp340 also recognizes the same motif in other bacterial proteins which exist in both Gram-negative and Gram-positive bacteria.

3.3.2. Method

Three different bacterial species were grown in different broth medium. Those species were *Streptococcus pyogenes, Streptococcus gordonii* and *Escherichia coli*. Human gp340 protein were purified from freshly collected, pooled saliva from six donors using gel filtration method. That followed by measuring the adhesion of *Streptococcus pyogenes* to gp340 protein using hydroxyapatite beads. They also used mass spectrometry for adhesion identification. That followed by an insertional inactivation of Spy0843 into *Streptococcus pyogenes*. The next step was binding of Spy0843-deficient mutants to purified gp340 from human saliva. That was followed by cloning and Expression of Spy0843 Recombinant Protein, named r0843.

ELISA technique was utilized to test the inhibition of binding between r0843 to gp340 by monosaccharides. Then the binding of truncated r0843 and other Lrr proteins to gp340 was tested. In the end, the effect of SRCR1A peptide on inhibiting the binding between gp340 and Lrr proteins was tested.

3.3.3. Main findings

The binding between gp340 protein which was isolated from human saliva and different clinical *Streptococcus pyogenes* isolates happened to varying degrees. Some bacterial strains were bound to Bovine serum albumin (BSA) coated beads, that might be caused by cross

reactions between bacteria and BSA. Despite of that, it was proven that all bacterial strains which were used in this experiment bound more efficiently to gp340 than to BSA.

Using mass spectrometry, they identified Spy0843 as gp340-binding protein. Sequence analysis showed a protein which composed of a 23-amino acid signal peptide, 143-amino acid N-terminal, a middle part of about 500 amino acids with 15 Lrr motifs and 360-amino acid C-terminus with an LPRTG cell wall-anchoring motif (illustrated in Figure 1a).

When the binding was tested of recombinant Spy0843 protein, r0843 protein and its truncated versions to gp340 protein, they found the full size r0843 (117 kDa) bound to gp340, while the truncated version that containing only the N-terminal (17 kDa) and the other truncated version that containing only C-terminal (38 kDa) did not bind to gp340. In the meantime, the truncated version that contains N-terminal together with the middle part of region that contains Lrr-motifs (63 kDa) was observed binding to gp340. From this the researchers concluded that the middle region of Spy0843 which contains Lrr-motifs is needed for gp340 binding.

When the specificity of r0843 binding to gp340 was investigated, using glycoproteins for control, the r0843 bound specifically to the same high molecular weight protein as the gp340-monoclonal antibody and not to other salivary glycoprotein receptors. This binding was observed to be calcium dependent since it was only observed in the presence of calcium chloride.

The microbial interaction with gp340 (binding between r0843 and gp340) was inhibited by SRCR1A peptide which contains -VEVL- amino acid sequence.

Three additional bacterial Lrr proteins were evaluated for binding to gp340. Those proteins are LrrG from *Streptococcus agalactiae,* BspA from *Tannerella forsythia* and YopM from *Yersinia enterocolitica.* A truncated form of LrrG which was only containing Lrr-region of the protein was also used. The LrrG and its truncated form which only contains Lrr-region was bound to gp340. The binding to gp340 was also observed with BspA proteins. The only exception was YopM protein which did not bind to gp340. This was explained as the YopM protein is a representative of a different Lrr-type which is shorter and contains an arrangement of leucines which is different from the arrangements in Spy0843.

3.4. Review of the third article

The title of the third article is "*Staphylococcus aureus* SasA is Responsible for Binding to the Salivary Agglutinin gp340, Derived from Human Saliva". It was written by Kenji Kukita, Miki Kawada-Matsuo, Takahiko Oho, Mami Nagatomo, Yuichi Oogai, Masahito Hashimoto, Yasuo Suda, Takuo Tanaka and Hitoshi Komatsuzawaa and was published in American Society for Microbiology Journal "Infection and Immunity", June 2013.

3.4.1. The purpose of the research

The authors of this article aimed at investigating the interaction between gram positive *Staphylococcus aureus* bacterium and saliva, and specifically to the gp340 protein receptor. They also wanted to identify the *Staphylococcus aureus* factor which is responsible for binding to gp340.

3.4.2. Method

First, different strains of *Staphylococcus aureus* were cultured and human saliva from a healthy volunteer was used to purify the gp340 protein. That followed by construction of srtA, sasA, and sdrE gene mutants of *Staphylococcus aureus*. Afterwards, recombinant SasA proteins were constructed which also included some truncated versions to be purified. Among those truncated recombinant proteins was rSasA-N with the N-terminal one-third, which retains the BR-domain (illustrated in Figure 2).

Then they verified those proteins using SDS-PAGE in a 7.5 % polyacrylamide gel, followed by silver staining. The following step was obtaining antiserum against rSasA-N protein by immunizing mice.

That followed by performing an assay for binding of *Staphylococcus aureus* to saliva and gp340-coated resin. In this part they used clarified saliva, after treatment of *Streptococcus mutans*, and purified gp340 at different concentrations.

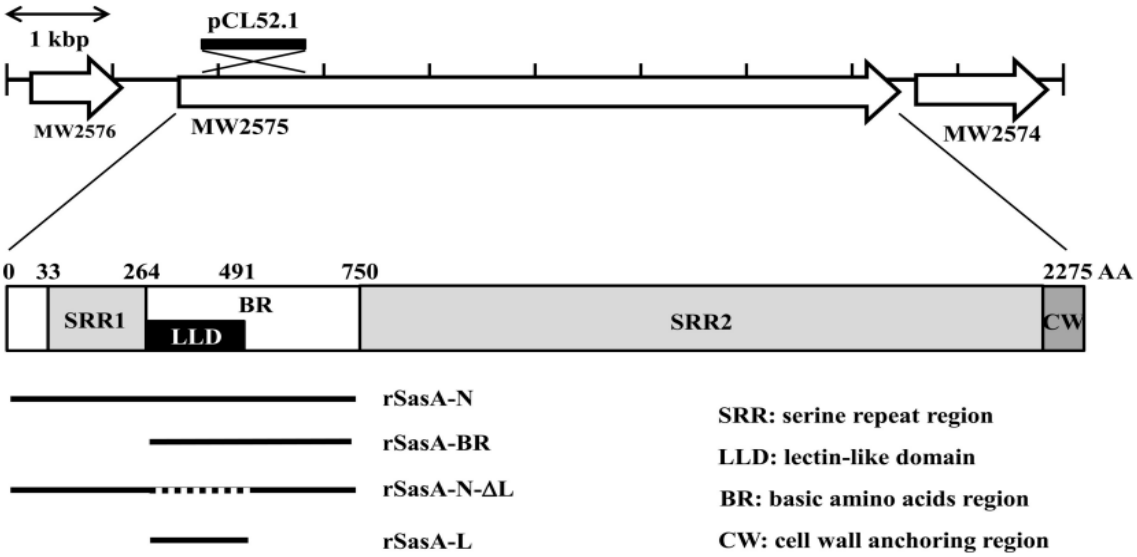


Figure 2:Showing a schematic representation of Staphylococcus aureus SasA protein. (Kukita et al., 2013)

In order to investigate the trypsin treatment impact on *Staphylococcus aureus* binding to gp340, the bacterium was treated with 0.5, 1, 2, or 4 µg/ml of trypsin for 10 min at 37°C. The bacterium was subjected to the binding assays after five washes with PBS. They investigated the inhibition monosaccharides, lectins and recombinant SasA protein on binding between *Staphylococcus aureus* and gp340 receptor.

Thereafter, they performed an assay for binding of recombinant SasA to gp340 using purified recombinant proteins which was obtained earlier. They investigated the inhibitory effects of mannose and N-acetylneuraminic acid on binding between rSasA-N protein to gp340,

They also performed a Surface plasmon resonance (SPR) analysis to study the molecular interactions between recombinant SasA-proteins and sugar. In the end they analysed the expression of SasA-protein in *Staphylococcus aureus* clinical isolates and they also investigated the binding between those clinical isolates and gp340-coated resin.

3.4.3. Main findings

As a result of investigating the binding of *Staphylococcus aureus* to saliva-coated resin, gp340-coated resin, BSA-coated resin, and noncoated resin, they found *Staphylococcus aureus* significantly bound to saliva and gp340-coated resin compared to BSA-coated resin

and noncoated resin. They also reported that *Staphylococcus aureus* did not bind to resin coated with saliva after treatment with *Streptococcus mutans* cells.

When it comes to the factors that affect the binding between *Staphylococcus aureus* and gp340, it was reported that treating *Staphylococcus aureus* with high doses of Trypsin reduced its binding efficiency to gp340 receptor. They also found that srtA knockout mutant of *Staphylococcus aureus* has lower efficiencies of binding to saliva- and gp340-coated resin. All sugars which was investigated as a possible factor that can affect the binding between *Staphylococcus aureus* and gp340 had only a weak binding-inhibitory affect except N-Acetylneuraminic acid which significantly inhibited the binding.

When it comes to the effect of recombinant SasA protein on *Staphylococcus aureus* binding to gp340, they failed to construct the full length rSasA protein of 228,4 kDa due to the large molecular mass, but they managed to construct a truncated rSasA-N with the N-terminal one-third which retains the BR domain. They found that the pre-treatment with rSasA-N inhibited the efficiency of *Staphylococcus aureus* binding to gp340-coated resin in a dose-dependent manner. Similar result obtained when using rSasA-BR protein which contain only BR-domain (illustrated in Figure 3). On the contrary, the inhibitory affect wasn't that strong when they used rSasA-L protein which contains only lectin-like domain in BR domain.

The recombinant rSasA-N protein was found to bind to gp340 in a dose dependent manner. The binding of rSasA-N to gp340 was inhibited to a bigger extent when N-acetylneuraminic acid was added to the assay. Mannose had little inhibitory affect while glucose have no effect on the binding at all. The SPR analysis showed that the recombinant rSasA-N bound significantly to NeuAcα (2-3) Galβ (1-4) GlcNAc sugar, while it didn't exhibit a significant binding with some other sugars like oligosaccharides that have glucose, galactose, mannose or GlcNAc as the terminal sugar. They also found that Maackia amurensis agglutinin (MAM) has a higher inhibitory effect on binding between *Staphylococcus aureus* and gp340 than the lectins Sambucus sieboldiana agglutinin (SSA).

When it comes to the expression of SasA levels in clinical isolates of *Staphylococcus aureus*, the quantitative PCR analysis showed variation in SasA between the different strains. Also, *Staphylococcus aureus* binding to gp340-coated resin varied among strains. They also identified a positive correlation between SasA expression in *Staphylococcus aureus* and binding ability to gp340 receptor which means that the bacterial strains with high expression of SasA protein have a high binding efficiency to gp340.

3.5. Review of the fourth article

The title of the fourth article is “Expression, purification and crystallization of the C-terminal LRR domain of *Streptococcus pyogenes* protein 0843”. It was written by Teemu Haikarainen, Vuokko Loimaranta, Carlos PrietoLopez, Pradeep Battula, Jukka Finnec and Anastassios C. Papageorgiou, and was published in the structural Biology Communications Journal "Acta Crystallographica Section F" 30th of April 2013.

3.5.1. The purpose of the research

The adhesion mechanisms in Gram-negative bacteria had been extensively investigated and understood while little was known about the same mechanism in Gram-positive bacteria, and this paper is trying to fill this gap.

To study the adhesion mechanism in Gram-positive bacteria, the authors of this paper chose to work with *Streptococcus pyogenes* and particularly with a protein known as Spy0843. It was earlier suggested that the Lrr-motif in Spy0834 takes part in the adhesion mechanism that helps bacteria to establish the host infection binding to gp340 proteins in saliva and epithelial cells (Loimaranta et al., 2009). The authors of this paper have initiated structural studies on the entire protein and its individual Lrr domains. In this study, the structure of one of two Lrr motifs in Spy0834 has been studied by X-ray crystallographic analysis. This to characterize common features in bacteria-gp340 interactions in general.

3.5.2. Method

They described the expression, purification, crystallization and structure data collection of the C-terminal Lrr domain of Spy0843 protein in *Streptococcus pyogenes* (illustrated in Figure 1a). The expression and purification of Lrr domain was done by PCR-amplification of the genomic DNA sequence which is encoding the C-terminal Lrr domain of Spy0843 protein. This DNA was ligated into a plasmid, which was then transformed into *Escherichia coli* for

protein expression and followed by purification of the Spy0843 protein. The initial crystallization was performed using the sitting-drop vapour diffusion method.

Structural data was collected from a single Spy0843 crystal which was soaked briefly in a reservoir solution. The crystal was afterwards placed in a stream of gaseous nitrogen at 100 K. The wavelength of X-ray used was 0,81230 Å and the data was collected in a form of 186 images. The distance between the detector and the crystal was set to 141 mm and the rotation range was adjusted to 0.3 °.

3.5.3. Main findings

When it comes to the expression and purification of Spy0843 protein, the SDS-PAGE gave a major band with molecular mass of about 33 kDa as it was expected, which prove that they managed to express and purify the correct protein.

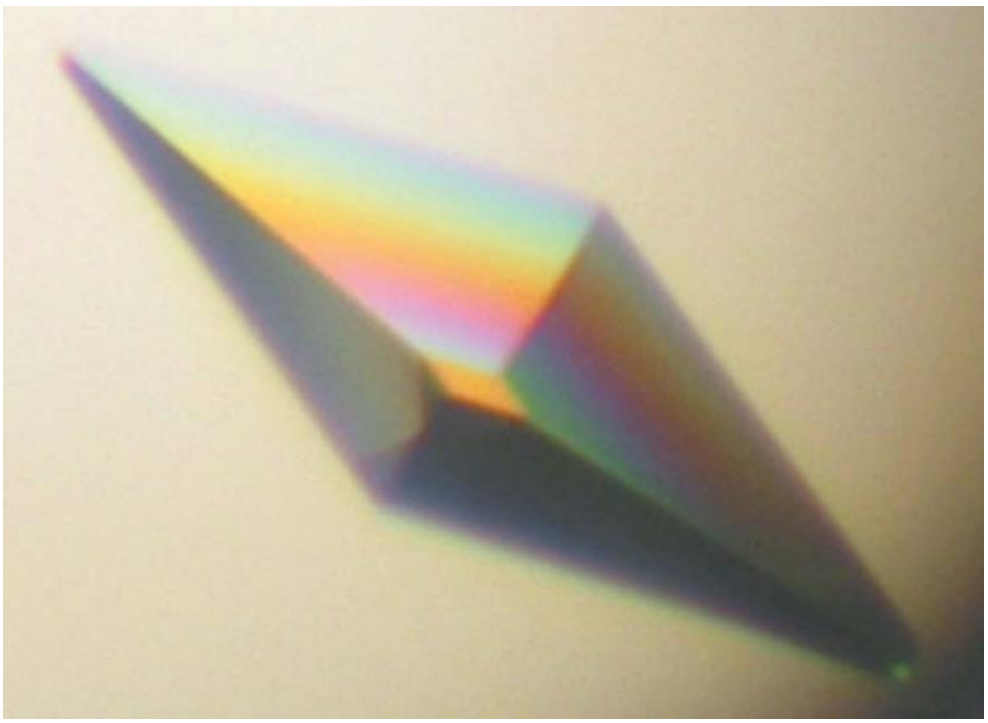


Figure 3:Showing a typical crystal of Spy-2 with approximate dimensions 0.35 x 0.15 x 0.15 mm. (Haikarainen et al., 2013)

When it comes to crystallization of recombinant Spy0843, it has been produced protein crystals with suitable size and the crystallization condition has been optimized (illustrated in Figure 3).

The structure determination of the C-terminal Lrr-domain of Spy0843 is achieved by single or multi wavelength anomalous dispersion method. The structure of this protein was not reported in the article, they only included the crystal structure which can be used in a further X-ray analysis.

3.6. Review of the fifth article

The title of the fifth article is “Structures of SALSA/DMBT1 SRCR domains reveal the conserved ligand-binding mechanism of the ancient SRCR fold”. It was written by Martin P Reichhardt, Vuokko Loimaranta,Susan M Lea and Steven Johnson and was published in Life Science Alliance, February 2020.

3.6.1. The purpose of the research

In this study, researchers worked with gp340 protein, in this paper named as Salivary Scavenger and Agglutinin (SALSA) protein which also known as DMBT1.

The purpose of this research is to provide a better understanding of the multiple ligand binding properties of gp340 by providing detailed information of its interaction surfaces. For this goal, they undertook an X-ray crystallographic study and provided an atomic resolution structures of SALSA SRCR domains 1 and 8.

3.6.2. Method

First, they carried out the expression of recombinant proteins by inserting the DNA of SALSA SRCR protein domains 1 and 8 were into a plasmid with a C-terminal His-tag which were transfected into an insert cell line (SR2). This was followed by protein expression, purification and crystallisation of the two SALSA SRCR domains. Crystallization was performed using the vapor diffusion method. Collecting the x-ray data from SRCR 1 and 8 domains of SALSA was carried out at 80 K temperature on beamlines 104, a wavelength of

1,0718 Å with a diamond light source. At the same temperature, the data was collected from SRCR8cat but this time at beamlines and a wavelength of 0,9762 Å.

The first ligand was Hydroxyapatite binding assay which is a phosphate-rich mineral essential for the binding of SALSA to the teeth surface, where it mediates antimicrobial effects. The second was Heparin binding assay which is a sulphated glycosaminoglycan as a mimic for the ECM/cell surface, for which binding of SALSA has been described to affect cellular differentiation and microbial colonisation. And the last ligand to be tested was Spy0843.

3.6.3. Main findings

SALSA includes 13 SRCR domains with high identity that ranges from 88 % to 100 % (illustrated in Figure 4). Each of those domains is composed of 109 amino acid residues where variations have been observed only in 9 of them.

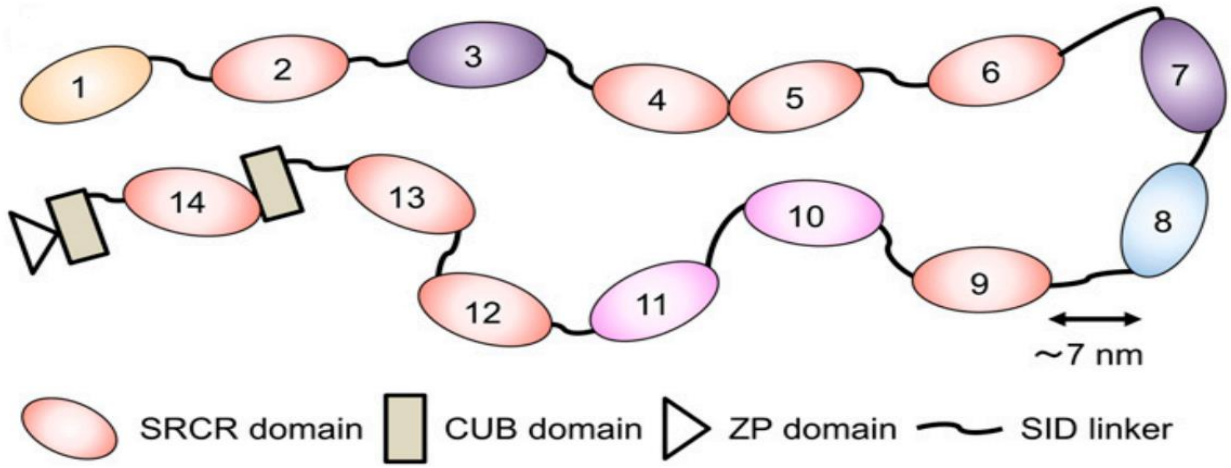


Figure 4:Showing a schematic representation of the domain organization of full-length SALSA/gp340 protein with SRCR1 and SRCR8 domains highlighted in green and blue, respectively. (Reichhardt et al., 2020)

The focus in this study was aimed at SRCR 1 and 8 domains since they are considered valid representations of all SALSA SRCR domains.

The crystallographic study of the two SALSA SRCR domains shows a classic globular SRCR-fold (illustrated in Figure 5), with four conserved disulphide bridges. The SRCR-fold contains an α-helix and one additional single helical turn. This fold is surrounded by N and C terminals which come together in a four-stranded β-sheet.

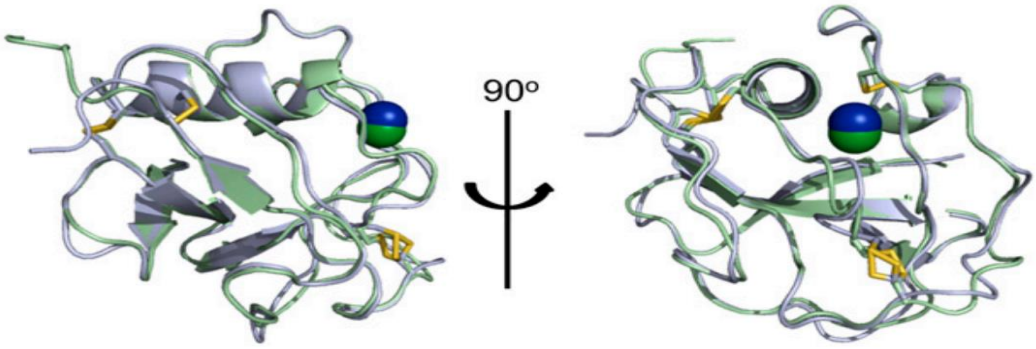


Figure 5:Showing the front and side views of an overlay of SRCR1 in green and SRCR8 in blue. It is also showing four conserved disulphide bridges in yellow. (Reichhardt et al., 2020)

Those domains are found to be stabilized by three Mg^{2+} ions buried in the globular folds (as illustrated in figures 5, 6 and 7).

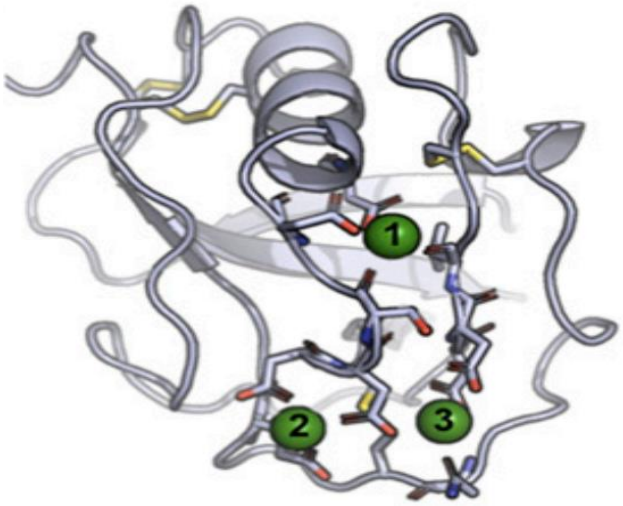


Figure 6:Showing a representation of the residues coordinating the three Mg2+ cations. (Reichhardt et al., 2020)

It was further verified that SALSA ligand binding interactions are calcium rather than magnesium dependent, by conducting a binding assay in an MgEDTA-containing buffer. The exchange of magnesium for calcium abolished ligand interactions, and that is a proof of SALSA binding calcium dependency.

When it comes to hydroxyapatite, heparin and Spy0843 binding assays, which performed to demonstrate the ability of SALSA to recognize and bind to a broad range of biological molecules, they find that the WT form of SRCR8 domain bound successively to all three ligands. In the meantime, the mutant form of SRCR8 domain with mutation affecting cation-binding site 2 didn't bound to those ligands.

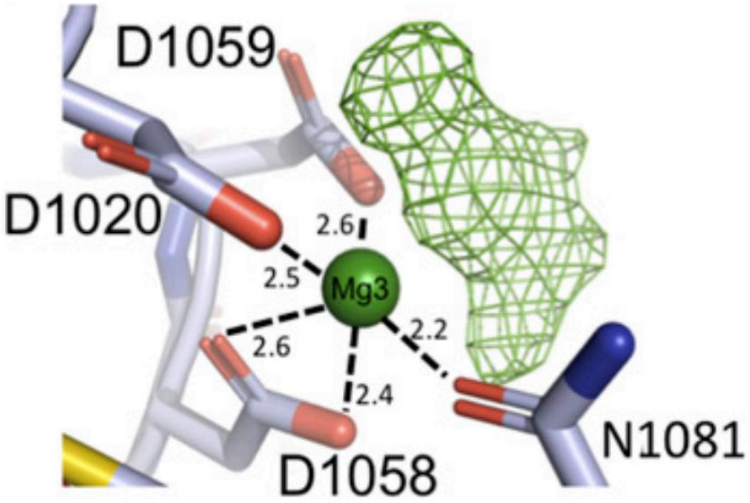


Figure 7:Showing a detailed view of the coordination of the Mg2+ cation at site 3. (Reichhardt et al., 2020)

This also happened with another mutant form of SRCR8 with mutation affecting both cation-binding sites 2 and 3. The reason is the fact that cation-binding sites are acting as bridges for ligand interaction, and when those sites get mutated they couldn't perform such function that effectively.

4. Discussion

The papers which were included in this review were collected by searching online databases and search engines for articles that deals with structure of Lrr-proteins in different bacterial species and how it interacts with gp340 receptors. The literature search was also highlighting the X-ray crystallography as a great macromolecule's structural determination method that can help us to study the protein structure at the atomic level.

After going through the search results, I chose five articles which were published over a period of 16 years, from 2004 until 2020, because that would give the reader of this paper a sense of how the research in this area has developed over the time. I also chose the articles considered to be most relevant to the main topic and which was written as a scientific research report.

When I analysed the articles, I emphasized the binding between Lrr-protein domains and gp340 receptor and the different factors that can have an inhibiting effect on this binding. Such factors can help the researchers to develop an anti-adhesion treatment for bacterial infections in the future.

Protein crystallization were used in two studies that was included in this paper. The study which was documented in the fourth article (Haikarainen et al., 2013) explained in details the vapour diffusion method which was used to crystallize the C-terminal Lrr domain of Spy0843 protein which is found in *Streptococcus pyogenes*. A similar method was also explained in the fifth article (Reichhardt et al., 2020) for crystallization of SRCR1 and SRCR8 domains of the gp340 (SALSA) protein.

In this study, the X-rays were used to determine the atomic structure of the crystallized proteins in order to study its molecular structure, the functions and interaction with other proteins. Both articles four and five are important because they give a detailed explanation of the method step by step in a way that can help other scientists to use it in future studies related to studying of macromolecules structures and functions.

According to the first article (Ikegami et al., 2004), there is a high similarity in amino acid sequence of Lrr motifs of LrrA protein that exists in *Treponema denticola* and other Lrr proteins in other bacterial species. They also found that Lrr-A protein plays a role in attachment of *Treponema denticola* to the epithelial cells. However, this study, which was made as early as 2004, didn't suggest which protein receptor that mediate such attachment in the epithelial cells.

This gap was filled with the study which was documented in the second article (Loimaranta et al., 2009) which found that the binding happened between Lrr motifs of Spy0843 protein in *Streptococcus pyogenes* and gp340 protein which was isolated from saliva. Similar findings were confirmed in the fifth article (Reichhardt et al., 2020), and also in third article (Kukita et al., 2013) which confirmed the binding between Staphylococcus aureus SasA-protein ang gp340 receptor without specifying which protein motif in SasA-protein which is responsible for the binding. This finding is important because it shows how the pathogenic bacteria bind to the host cell when it establishes the infection.

Different studies which were included in this paper investigated the possible factors that can inhibit the binding between the bacterial Lrr-proteins and gp340 receptor. The study which was documented in the second article (Loimaranta et al., 2009) showed that the binding between those two has been inhibited by the SRCR1A peptide which contains a -VEVL- amino acid sequence. Also, it was proved in the study that was documented in the third article (Kukita et al., 2013) that the binding of *Staphylococcus aureus* to salivary gp340 protein was

inhibited by N-Acetylneuraminic acid. Those findings are so important as they can be used in the future to develop anti-adhesion treatments that can replace antibiotics.

So far, we know which part of gp340 receptor that mediate the binding between the bacteria and host cells, but we lack the knowledge of the detailed molecular structure of Lrr-motif binding to gp340.

We also know which compounds that can stop this binding *in vitro*, but nothing of this have been tested *in vivo* to determine the actual affect of those eventual treatments on humans and whether it has side effects.

What we need to do further is to carry out X-ray crystallographic study of Spy0843 protein in order to obtain a detailed molecular structure of Lrr-motif binding to the gp340 receptor. Such detailed information is expected to provide more knowledge on the mechanisms of the gp340-specific binding to bacteria.

We also need more investigations on other possible inhibitory factors, in order to extend the treatment possibilities. Then we need to test the inhibitory factors that assumed to have the minimum side effects *in vivo*, to see if this would actually work.

5. References

Alberts, B. (2004). *Essential cell biology*. New York, NY: Garland Science Pub.

Frueh, D. P., Goodrich, A. C., Mishra, S. H., & Nichols, S. R. (2013). *NMR methods for structural studies of large monomeric and multimeric proteins*. Current opinion in structural biology, 23(5), 734–739.

Haikarainen, T., Loimaranta, V., Prieto-Lopez, C., Battula, P., Finne, J., & Papageorgiou, A. C. (2013). *Expression, purification and crystallization of the C-terminal LRR domain of Streptococcus pyogenes protein 0843*. *Acta crystallographica. Section F, Structural biology and crystallization communications, 69(Pt 5), 559–561.*

Hare R. (1982). *New light on the history of penicillin*. Medical history, 26(1), 1–24.

Ikegami, A., Honma, K., Sharma, A., & Kuramitsu, H. K. (2004). *Multiple functions of the leucine-rich repeat protein LrrA of Treponema denticola*. Infection and immunity, 72(8), 4619–4627.

Kobe, B., & Deisenhofer, J. (1994). *The leucine-rich repeat: a versatile binding motif.* Trends in biochemical sciences, 19(10), 415–421.

Kukita, K., Kawada-Matsuo, M., Oho, T., Nagatomo, M., Oogai, Y., Hashimoto, M., Suda, Y., Tanaka, T., & Komatsuzawa, H. (2013). *Staphylococcus aureus SasA is responsible for binding to the salivary agglutinin gp340, derived from human saliva*. Infection and immunity, 81(6), 1870–1879.

Lobanovska, M., & Pilla, G. (2017). *Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future?* The Yale journal of biology and medicine, 90(1), 135–145.

Loimaranta, V., Hytönen, J., Pulliainen, A. T., Sharma, A., Tenovuo, J., Strömberg, N., & Finne, J. (2009). *Leucine-rich repeats of bacterial surface proteins serve as common pattern recognition motifs of human scavenger receptor gp340*. The Journal of biological chemistry, 284(28), 18614–18623.

Papageorgiou, A. C., & Mattsson, J. (2014). *Protein structure validation and analysis with X-ray crystallography.* Methods in molecular biology (Clifton, N.J.), 1129, 397–421.

Reichhardt, M. P., Loimaranta, V., Lea, S. M., & Johnson, S. (2020). *Structures of SALSA/DMBT1 SRCR domains reveal the conserved ligand-binding mechanism of the ancient SRCR fold*. Life science alliance, 3(4), e201900502.

Wang, H. W., & Wang, J. W. (2017). *How cryo-electron microscopy and X-ray crystallography complement each other.* Protein science: a publication of the Protein Society, 26(1), 32–39.

Wüthrich K. (2003). *NMR studies of structure and function of biological macromolecules (Nobel Lecture)*. *Journal of biomolecular NMR*, *27*(1), 13–39.