HIRSZFELD INSTITUTE OF IMMUNOLOGY AND EXPERIMENTAL THERAPY, POLISH ACADEMY OF SCIENCES

Doctoral dissertation

Immune epitopes of pneumococcal endolysins Cpl-1 and Pal and their role in shaping immunoreactivity of the enzymes

mgr inż. Marek A. Harhala

supervisor: prof. dr hab. Krystyna Dąbrowska

Wrocław, 2022

Epitopy endolizyn pneumokokowych i ich rola w tworzeniu odpowiedzi odpornościowej na te enzymy

1. Table of Contents

2. List of Figures

3. List of Tables

4. List of abbreviations

bp – base pairs (DNA)

B834 - expression strain of *Escherichia coli*

BLAST - Basic Local Alignment Search Tool

C57BL6 – inbred strain of mice used for laboratory purposes

CDC -. Centers for Disease Control and Prevention

CFU - colony-forming unit

Da Dalton - atomic mass unit

DH5α szczep *E. coli* DH5α

DNA - deoxyribonucleic acid

DNA-se deoxyribonuclease

DSSP - Dictionary of Secondary Structure of Proteins

E. coli – Escherichia coli

ELISA - enzyme-linked immunosorbent assay

FDA – U.S. Food and Drug Administration

IgG immunoglobulin G

HIEET Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences

IL interleukin

LB – Luria-Bertani

LPS - lipopolysaccharide

MDR – multi-drug resistant

NCBI ang. National Center for Biotechnology Information

OD_{600} - optical density measured at 600nm

ORF - open reading frame

PAN Polska Akademia Nauk

PBS - phosphate buffered saline

PCR - polymerase chain reaction

PMSF Phenylmethylsulfonyl fluoride

RNAse - ribonuclease

rRNA – ribosomal ribonucleic acid

SD - standard deviation

SDS–PAGE - sodium dodecyl sulfate – poliacrylamide gel electrophoresis

SPF - specific pathogen free

Tris tris(hydroxymethyl)aminomethane

USD – United States Dollar

WHO – World Health Organization

5. List of publications

Results presented in this dissertation are published in articles listed below.

Harhala M, Nelson DC, Miernikiewicz P, Heselpoth RD, Brzezicka B, Majewska J, Linden SB, Shang X, Szymczak A, Lecion D, Marek-Bukowiec K. *Safety studies of pneumococcal endolysins Cpl-1 and Pal*. Viruses. 2018 Nov 15;10(11):638

Harhala M, Gembara K, Miernikiewicz P, Owczarek B, Kaźmierczak Z, Majewska J, Nelson DC, Dąbrowska K. *DNA Dye Sytox Green in Detection of Bacteriolytic Activity: High Speed, Precision and Sensitivity Demonstrated With Endolysins*. Frontiers in microbiology. 2021:3129

Harhala M, Gembara K, Nelson DC, Miernikiewicz P, Dąbrowska K. *Immunogenicity of Endolysin PlyC*. Antibiotics. 2022 Jul 18;11(7):966

6. Introduction

1. Proteins as potent medical tools

The use of various substances to solve medical and technological problems and suiting these substances to our needs is one of the most potent tools for humankind. These substances can be acquired from natural sources, including biological ones. They can be produced by virtually all kinds of living cells, controlled by natural machinery employing genetic code as the source of information and protein synthesis system. Members of this group of useful biological products are proteins, proved to be extremely diverse and to represent a myriad of different activities. This variety can be observed within the entire biome on Earth, consisting of at least 8.7 million species (by low estimates) (1). Structural analysis revealed that simple elements, such as proteins, create the complexity of the entire biome. Deep analysis of this biome led to the development of ideas regarding levels of organisation, like integrative levels in biology rooted in broader philosophical trends (2,3). This idea presents a process in which smaller elements at lower levels of organisation (e.g. proteins) create higher levels of organisation (e.g. cells) which reciprocally affect lower levels, for instance by creating a controlled environment inside the cells, affecting them and allowing for new synergies and properties to emerge, often called in short emergent properties. These interactions at the level of the biome create incredibly complicated materials and substances with a plethora of purposes, from soft bodies of medusas to hard bone structures or turtle shells.

Interest in exploitation of this potential of the biological world increased with the development of modern techniques such as protein expression in many expression systems (molecular biology) or molecular modelling (bioinformatics). The low cost of *de novo* synthesis of DNA coding for proteins that can be expressed in microbial systems revolutionised biochemistry and allowed for creation of almost any recombinant protein and for testing their functions and applications in relatively simple laboratories (4,5). This created however the biggest challenge for protein research: how to

choose an optimal protein sequence that can be expressed, purified, and used to achieve specific, pre-defined functionality. This challenge is currently difficult due to two reasons:

- limited knowledge necessary for accurate calculations of quantum interactions crucial for protein functionalities
- the very high number of possible protein sequences (resulting from combinations of available amino acidic constituents).

The first problem relates to challenges in prediction of interactions at the atomic resolution. These are crucial for determination of protein folding, interactions with other molecules, and chemical reactions, often being the most important manifestation of a protein function. Precise calculations could allow us to establish general, simplified rules that govern proteins' synthesis, folding, and interactions. Lack of this knowledge forces scientists to perform extensive wet-lab research and a trial-and-error approach to understand functions of a specific protein. Still, this knowledge cannot be extrapolated beyond similar proteins and application, all other holistic conclusions being greatly impeded. This challenge is shared with chemical sciences and nanotechnology where detailed knowledge about the structure-activity relationship or chemical activity would speed up development of new chemicals and increase its potential to solve problems (4–6).

The second problem is connected with the number of possible protein sequences, which is incredibly large. In order to understand this problem, we need to consider a simplified theoretical model of a protein, which is a chain made of amino acids (primary structure) that interact with each other in three-dimensional space causing folding (secondary and tertiary structure), creating a macromolecule capable of interaction with other proteins/macromolecules through its entire surface (quaternary structure). This macromolecule can by subjected to chemical modification (post-translational

modifications) and it is able to interact with virtually any molecule in the environment. The primary structure of a protein, a sequence of amino acids on the protein backbone, is the main source of protein diversity. We can calculate the number of potentially existing proteins by limiting our consideration only to 20 natural amino acids ubiquitously coded in the a genome (proteinogenic amino acids). Even for very short oligopeptides, the number of combinations is very large and increases rapidly with protein length; for instance, there are 8 000 (20^3) possible combinations of tripeptides, and 3 200 000 (20^5) of possible pentapeptides in total. Proteins do not have an intrinsic limit of their length and naturally occurring proteins can be thousands of amino acids in length. A median length for a human proteome equals to 375 aa allowing for $7.7*10^{487}$ different amino acid sequences (7). By comparing this with the number of atoms in the universe (10^{82}) it can easily be concluded that there are far too few atoms in the observable universe to synthesise even a tiny fraction of the theoretically possible protein combinations (8–11).

Secondary structure describes spatial configuration of amino acids close to each other in the sequence of amino acids. This structure strongly depends on angles in the peptide bond since its steric hindrances strongly limit possible configurations (12,13). There are 8 secondary structures originally recognised in the Dictionary of Secondary Structure of Proteins (DSSP) and commonly used today, namely: alpha helices, 3.10 helix (3-helix), pi-helix (5 helix), beta strand, bridge (a residue in isolated beta-bridge), turn, bend and a loop (irregular) (14,15).

Tertiary structure describes the spatial configuration of amino acids distant on the protein backbone. From the perspective of protein modelling this structure is complicated to predict due to virtually infinite possibilities and currently available protocols fail to provide well-defined effects (16). The most typical approach *in silico* is in fact finding another protein with a detectable homology that has been physically analysed (crystallography) and to build on similarities between this already solved

protein structure and the new one under investigation. Another approach is to find energetically favoured spatial configurations (local energy minimum), however large proteins tend to have many low-energy tertiary structures that can be stable in biological conditions, but difficult to predict *in silico*. Prions causing diseases are a stark example of how differences in protein folding and different stable local energy minima in conformational space can impact the function of proteins.

Quaternary structure is a term used to describe interactions and spatial configuration of macromolecules (peptides) interacting with each other usually forming multi-peptide complexes that cooperate with each other. To make these interactions happen, a part of the protein structure needs to fit the other macro-molecule, both creating stronger-than-thermal energy interaction. Typically, it is a sum of van der Waals, hydrogen bonds, and interaction between charged particles on the surface of molecules. Often such interactions are characterized by specific distribution of electric charges on the surface or specific shape of the molecule. After creating many relatively weak interactions a molecule can create stable interactions with other, different macro-molecules in multiple places of its surface.

Levels of protein organisation as discussed herein are commonly accepted but not the only possible ones. Other structures such as super-secondary structures, motifs and domains are also applicable (17). One other possible approach in protein science is to focus on functional units and to try to overcome the complications originating from combinatorics, thus simplifying rules for protein *de novo* modelling. Functional classification of protein subunits in sets that serve specific function is increasingly often used. This classification allows protein structures to be grouped in clusters from the smallest, spanning one amino acid (such as post-translational modifications) through motifs and domains, up to the largest, multi-protein complexes (18).

For instance, the Pfam database (version 34.0 at http://pfam.xfam.org/) represents currently almost 20 000 families and 645 classes of protein domains. In the Eukaryotic Linear Motif relational database storing experimentally validated short linear motifs (http://elm.eu.org/) there are, at this moment, 3 934 experimentally validated instances (including post-translational modification sites) categorized in 317 classes.

As a reference, the human genome codes for approximately 3.5×10^4 proteins. Therefore, almost no proteins are synthesized in comparison to the theoretically possible $7.7*10^{487}$ amino acid combinations for a 375 aa long protein (median length of the human proteome) (19,20). 22 000 of the investigated human proteins contain around 35 000 instances of folded domains belonging to 10 000 families, and according to estimations there are 10^5-10^6 functional instances (post-translational modifications, motifs, domains, complexes) in total in the human proteome (18).

The number of possible quaternary structures of dimers is estimated to be about 10^4 times larger than measured, so an experimentally confirmed database of quaternary structures is not likely to be close to completion in the near future (21). Further complications arise when we consider non-proteogenic amino acids (non-ribosomal peptides).

The above problems result in a lack of general rules that allow for quick and efficient protein design for specific requirements. Protein research is currently dependent on proteins found in nature and on further modifying these proteins to possibly achieve a desired outcome.

In conclusion, our possibilities to exploit the potency of protein design and analysis are still very limited. An efficient approach to protein modelling needs to be goal-driven and it still requires tools that make it possible to design a protein/macromolecule that meets specific requirements.

2. Lysins as potential antibacterial biological drugs

Protein drugs are a rapidly growing part of the health industry. Currently there are 704 protein drugs out of 4196 total drugs approved in the DrugBank Online database. Also, of all the drugs listed in the FDA-curated THPdb drug database 239 there is approved proteins and peptides as well as 380 approved variants of these proteins (22,23). At the same time, development of new (not yet registered) protein drugs is probably the most dynamic area of protein research. Lysins are proteins with enzymatic activity, able to destroy the bacterial cell wall. Endolysins are a group of phage-encoded lysins (phage lysins) (24,25) often called enzybiotics. Enzybiotics is a term coined by Nelson, Loomis and Fischetti to describe antibacterial enzymes; "phage lysin" was mentioned as a specific example of this group of proteins that have been proposed as highly applicable in medicine (26). Particularly, enzybiotics are proposed as a response to increasing multi-drug resistance in bacteria. Enzybiotics have a potential to address problems with development of new antibacterials, since they are very different to traditional antibiotics, thus being unaffected by cross-resistance (26). The WHO has classified endolysins as antibacterial agents under development – 'Hence, all these products can in principle be considered innovative, as they target new structures through new modes of action. So far, these non-traditional agents have been developed for pre-emptive or adjunctive therapy'(27).

The total number of naturally existing endolysins is not known, but 2629 putative endolysins were discovered in samples of uncultured phage genomes only in one study published in 2018, demonstrating that large amounts of phage endolysins are yet to be discovered (28). Endolysins are characterised by a fast mode of action, and the high specificity of their bactericidal activity is necessary

for evolutionarily successful phage strains, while on the other hand it protects the majority of the bacterial microbiome from destruction when an endolysin is used *in vivo*.

3. Bacteria as a substrate

Bacteria are one of the most basic units of biological diversity in the three-domain system and describe unicellular organisms without a nuclear membrane and consistent rRNA differentiating them from Archaea (29,30). The bacterial cell wall that is targeted by endolysins is chemically very distinct from any structures that can be found in animal or human bodies. One of its main functions is to stabilize the cell membrane withstanding high turgor pressure pushing outwards (31). Chemically, the bacterial cell wall is made of peptidoglycan: built from long glycan strands consisting of alternating N-acetylglucosamine and N-acetylmuramic acid residues, cross-linked with short peptides, most often with L-Ala-g-D-Glu-meso- A_2pm (or L-Lys)-D-Ala-D-Ala (A_2pm, 2,6-diaminopimelic acid) (31). This feature has been observed in the peptidoglycan of all bacterial groups to date. Also, features such as presence of a characteristic sugar (MurNAc), of G-bonded D-Glu, of L–D (and even D–D) bonds and of nonproteinogenic amino acids (like A_2pm) are considered distinctive for the bacterial cell wall (31). Peptidoglycan often has attachments (bound covalently or not) that physically alter the surface of the cell wall. The most common include teichoic acids – functionally related anionic glycopolymer (31–33). This structure, highly conservative among bacteria, is the target for endolysins.

Endolysins destroy covalent bonds holding the cell wall solid and they create holes that release the bacterial spheroplast, resulting in disintegration of entire cells (34,35).

Endolysin are categorized based on the type of chemical bond they lyse (36):

1. endopeptidases – targeting peptide bonds including L-alanoyl-D-glutamate endopeptidase and interpeptide bridge endopeptidase;
2. amidases targeting amide bonds, including the N-acetylmuramoyl-L-alanine amidases;
3. glycosidases targeting glyosidic bonds including N-acetyl-β- D-glucosaminidases;
4. 4N-acetyl-β-D-muramidases;
5. transglycosylases (37).

4. Endolysin Cpl-1 and Pal

Pal lysin (UniProt ID: ALYS_BPDP1) is classified as 3.5.1.28 N-acetylmuramoyl-L-alanine amidase, since it acts on the carbon-nitrogen non-protein bond in linear amides. This enzyme is 196 aa long and it was described as early as 1983. It requires choline in the teichoic acid of the cell wall for its full activity, at the time being the first pneumococcal phage-encoded lysin to be isolated (38). Three amino acids (34, 99, 111) are crucial for Pal activity. The first 142 amino acids are considered the 'active domain' followed by repeats of the 'cell wall binding domain'.

Cpl-1 (UniProt ID: LYS_BPCP1) is classified as 3.2.1.17 (lysozyme in glycosidases). Its structure was published in 2003 using the single-wavelength anomalous diffraction method (39). Residues 1-188 create a catalytic (active) domain folding into an irregular $(\beta/\alpha)_5\ \beta_3$ (40). Further residues form a linker (189-199 aa), and CI and CII domains. CI and CII domains contain choline binding sites. CI is divided into repetitive structures designated p1-p4 and CII contains p5, p6, and the C-terminal tail (324-339 aa). Comparisons of Cpl-1 structure to some better known complexes of chitinolytic enzymes and substrates indicated that peptidoglycan (substrate) is placed along a negatively charged groove formed by its catalytic domain (41).

Both Pal and Cpl-1 enzymes can kill multiple streptococcus strains and they destroy a vast majority of bacterial cells in a sample within a few minutes after their contact with the substrates, even in relatively low concentrations of 10 mg/L (42).

5. *Streptococcus pneumoniae* as a difficult bacterial pathogen in humans

Throughout the human body there are many ecological niches for bacteria to colonize, constituting the microbiome. Microbiomes have been under intensive investigations in the last 20 years and have been proved to have a significant impact on human health, for instance to have immunoregulatory effects, to influence severity of drugs side effects, or to affect pharmacokinetics of pharmaceuticals (43). Bacteria are known to develop specific interactions to affect organisms of their hosts, such as *Bacteroides fragilis* using the capsular polysaccharide A to interact with plasmacytoid dendritic cells, effectively promoting IL-10 secretion from CD41 T cells (44).

However, some elements of microbiomes may have detrimental effects on health conditions, particularly when they outweigh other, beneficial microbes and when they propagate in large amounts. Even in modern times after almost an age of successful antibiotic therapies, according to the WHO, lower tract respiratory infections are the 4th worldwide, 2nd and 6th in low-income and high-income countries respectively, most common cause of death(45). Limitations of antibiotic use have been known for a long time; even Alexander Fleming in his Nobel Prize speech talked about an age of rising microbial resistance. In the last decades, new strands of multi-drug resistant bacteria have emerged, with some of strains resistant to all known antibiotic agents in medical use (46). This caused the WHO to repeatedly call for development of new antibiotics with new modes of action for a little more than two decades (27,47–50). Multi-drug resistant bacteria (MDR) are most com-

monly, repeatedly reported to increase lengths of stays in hospital, increasing mortality, and rising general costs of treatment between 3000 and 40 000 USD per patient (51).

Streptococcus pneumoniae is a potentially difficult bacterial pathogen. It is a lancet-shaped, Gram-positive bacterium, commonly inhabiting the respiratory tract. These bacteria are facultative anaerobes and they are relatively common in humans: 20-60% of children and 5-10% of adults without children are colonized by this bacterium (52). *Streptococcus pneumoniae* may cause difficult and even life-threating infections. Pneumonia, bacteraemia and meningitis are commonly recognized clinical syndromes of invasive pneumococci.

New, emerging antimicrobial therapies that affect bacteria in a different manner than traditional antibiotics are continuously under development. Nelson, Loomis and Fischetti proposed and demonstrated the applicability of endolysins in treatment of bacterial infections in mice *in vivo* in 2001 (26,35). Since then, use of extremely active phage endolysins has remained one of the possible treatments alternative to traditional antibiotics. Phage therapy focuses on the use of bacterial viruses (phages) that when administered to a patient can infect and destroy harmful bacteria, at the same time amplifying themselves and circulating in an organism, preventing reinfection (53,54). However, endolysins seem to lyse bacteria more rapidly, being less prone to filtering and removal by the mammalian immune system than whole phage virions. They also do not contain any genetic information and they are not able to mutate, thus being better controlled. In summary, active destruction of vital, typically bacterial structures, controllable character and extraordinary activity are strong supporting factors in choosing endolysins for viable medical and veterinary treatments in future.

6. Immune response as the challenge in application of protein drugs

Advantages of enzymes that could be used in medical therapies are described as "(…)superb specificity, well tolerated by the body, effective treatment, and versatility to modification and development"(55). Successful clinical trials of human insulin in 1982 marked the beginning of modern use of recombinant proteins in medicine and proved the potential of this type of drugs. One of the most important disadvantages of protein-based drugs is however connected to the proteinaceous nature of enzymes. Proteins are efficient inducers of a specific immune response that is aimed at recognition of foreign proteins. For human-derived protein drugs, such as insulin, the immune response can be much less pronounced, but microbial, viral, and newly designed (recombinant) proteins tend to induce a strong specific response from the immune system (56). Thus, the one of major challenges in medical applications of enzymes, including those of antibacterial activity, is to minimize the possibility that the enzyme will be efficiently neutralized by specific antibodies, induced by the treatment.

The major neutralizing agent produced by a specific part of the immune system is immunoglobulin G (IgG). It is capable of specific recognition of foreign proteins, accounting for 10-20% of proteins in human plasma as the major class of all immunoglobulins. The role of IgG is to recognise specific parts of antigens and trigger an immune response against them. As a result, foreign proteins are neutralized, thus limiting the efficiency of non-human enzymes in medical use. This means that specially repeated treatments with the same protein drugs are highly likely to fail due to the specific immune response and a significant level of IgG targeting the drug, as a result of the first round of treatment (55,57,58).

7. The aim of the study

Main goal of this work is to create variants of endolysins Cpl-1 and Pal with improved performance *in vivo,* particularly less prone to neutralization by specific antibodies without losing bacteriolytic activity. This will be done by completion of five steps:

1. Safety assessment of Cpl-1 and Pal.
2. Developing a new method for precise analysis of bacteriolytic activity of endolysin variants.
3. Determination of amino acids within Cpl-1 and Pal structure that mediate interactions of these endolysins with the immune system interacting with IgG antibodies.
4. Bioinformatic analysis and design of new variants of endolysins.
5. Expression and analysis of bacteriolytic activity and immunogenicity of selected Cpl-1 and Pal variants.

8. Materials and Methods

1. General materials

Genes coding for the investigated endolysins:

1. Cpl-1 wild type (WT): a lysozyme (acc. no.: CAA87744.1) from Streptococcus phage Cp-1 (acc. no. Z47794.1).
2. Pal wild type (WT): endolysin (acc. no.: YP_004306947.1) from Streptococcus phage Dp-1 (acc. no.: NC_015274.1).
3. PlyCA wild type (WT): PlyCA (acc. no.: NP_852017.2) from Streptococcus phage C1 (acc. no.: NC_004814.1)

4. PlyCB wild type (W): PlyCB (acc. no.: NP_852015.1) from Streptococcus phage C1 (acc. no.: NC_004814.1).

Expression vectors:

1. pBAD24 expression vector for *E. coli* – obtained with cloned ORFs from prof. Daniel Nelson, Institute for Bioscience & Biotechnology Research (IBBR) (Univeristy of Maryland, MD, USA)
2. pBAD/His_A expression vector for *E. coli* – obtained with cloned ORFs from prof. Daniel Nelson, Institute for Bioscience & Biotechnology Research (IBBR) (Univeristy of Maryland, MD, USA)

Bacteria strains:

1. DH5α: *Escherichia coli* DH5α (Invitrogen) genotype: F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1 (used for plasmid storage)
2. B834: *Escherichia coli* B834 (DE3) (OverExpress) genotyp: F- *ompT gal hsdSB* (rB-mB-) *met dcm Ion* λDE3 (used for protein expression)
3. Streptococcus pneumoniae strain PN135 (strain from Polish Collection of Microbes) susceptible to optochin and endolysins.

Growth media:

a) LB Broth high salt, pH = 7,5 (Fluka):

- casein enzymatic extract 10 g/L,
- yeast extract 5 g/L,

- NaCl 10 g/L,
- 2% agar was added if necessary for solid medium (BioMerieux).

b) THY broth (Todd-Hewitt broth with 2% Yeast extract) (Sigma)

- beef heart infusion 500 g/L,
- peptic digest of animal tissue 20 g/L,
- dextrose 2 g/L
- NaCl 2 g/L
- Na_3PO_4 0.4 g/L,
- Na_2CO_3 2.5 g/L,
- yeast extract 20 g/L

c) Liquid medium for eukaryotic cells (Laboratory of General Chemistry HIEET)

- 10% bovine serum
- 4,5% glucose
- 1% aminoacid
- 1:1 OR+ Gmax
- 1 mM pyruvate
- streptomycin 100 μg/mL
- ampicillin 100 μg/mL

1. Animals

We've used C57BL6 mice, (Mossakowski Medical Research Institute, Polish Academy of Sciences), 6-12 weeks old, male or female. Animals were held in specific pathogen free conditions (SPF) prior to experiments.

2. Equipment

- fridge-freezer (Hoover A Class)
- fluorescence plate reader (EnSpire Multilabel Plate Reader Perkin Elmer)
- CO_2 incubator CO14S (New Brunswick Scientific)
- universal incubation shaker Minitron (Infors HT)
- universal incubation shaker Unitron (Infors HT)
- laminar flow hood Biohazard II mini (POLON Poznań S.A.)
- laminar flow hood Biohazard II (Esco)
- G:BOX ge documentation system (Syngene)
- pH indicator (Fe-20 FiveEasy TM pH, Mettler Toledo)
- magnetic stirrer (IKA RTC Basic)
- microscope CX31 (Olympus)
- single- and multi-channel pipettes (Eppendorf)
- Spektrophotometer (BioPhotometer Eppendorf)
- Thermocycler (Eppendorf)
- Vortex (IKA MS 3 Basic)
- laboratory scale (Scout Pro 6000 g OHAUS)
- centrifuge (Mini Spin Plus Eppendorf)
- centrifuge (Sigma Laboratory Centrifuges 6-16K)
- ultra-low temperature freezer (Innova U101 New Brunswick Scientific)
- vertical electrophoresis system (Cleaver Scientific Ltd)
- horizontal electrophoresis system (Bio-Rad)

3. Ethical statement

All the human subjects gave their informed written consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Bioethics Committee of Regional Specialist Hospital in Wrocław (Project identification code: KB/nr 2/2017).

All animal experiments were performed according to EU directive 2010/63/EU for animal experiments and were approved by the 1st Local Committee for Experiments with the Use of Laboratory Animals, Wrocław, Poland (project no. 76/2011). The authors followed the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines.

4. Protein expression

Expression vectors with protein sequences coding Cpl-1, Pal and PlyC (PlyCA and PlyCB) proteins were originally acquired from the laboratory of prof. Nelson (Institute for Bioscience and Biotechnology research, University of Maryland) on pBAD24 plasmids with C-terminal 6×-His tag. For creation of Pal variants, new variants were synthesized *de novo* commercially with similar C-terminal 6×-His tag (BioCat GmbH) and cloned into pBAD/His_A vector (ThermoFisher Scientific). As for Cpl-1 variants, Cpl-1 WT and Cpl-1 v4 were synthesized and cloned into pBAD/His_A expression vector (BioCat GmbH). Variants were created by PCR Phusion Blood Direct Kit (ThermoFisher Scientific) with use of forward primers starting in the place of modification and reverse complement sequence starting at a first base before modification start served as a REV primer. Primers were 40bp in length optimized for the same temperature. Plasmid sequences were later confirmed by Sanger sequencing. Vectors were expressed in E. coli B834(DE3) cells (EMD), and grown at 37 °C with shaking in Luria-Bertani (LB) broth and ampicillin (100 μg/mL) (both from Sigma-Aldrich), until OD_{600} reached 1.0. Then, a protein expression was induced by the addition of

arabinose at a final concentration of 0.25%. The culture was subsequently incubated overnight at 20 °C with aeration by shaking.

5. Protein purification

Bacteria were harvested using centrifugation and suspended in PBS (Na_2HPO_4 0,91738 mg/L, KCl 200,0 mg/L, KH_2PO_4 200,0 mg/L, NaCl 8,0 g/L, $Na_2HPO_4x12H_2O$ 2,9 mg/L, pH 7.2), which was supplemented with PMSF (1 mM) and lysozyme (0.5 mg/mL). The slurry was incubated for 6–7 h on ice and lysed using the freeze-thaw method. Mg^{2+} (up to 0.25 mM), DNAse (up to 20 μg/mL), and RNAse (up to 40 μg/mL) were then added to the extract and allowed to incubate on ice for an additional 3 h. The fractions were separated using centrifugation (12,000× g, 30 min, 4 °C) and the soluble fraction (supernatant) was collected.

For experiments not including endolysin variants proteins, samples were then incubated with NiN-TA agarose (Qiagen) at room temperature washed with PBS (5× volume of the agarose) and an increasing concentration of imidazole (0 mM, 50 mM, 100 mM, 250 mM, and 500 mM). The 100 mM and 250 mM fractions containing the eluted endolysins were dialyzed against PBS at 4 °C. were separated using gel filtration (fast protein liquid chromatography) on a Superdex 75 10/300 GL column (GE Healthcare Life Sciences). The final step was LPS removal, which was performed with an EndoTrap Blue column (Hyglos GmbH, Munich, Germany). Purified, endotoxin-free protein samples were dialyzed against PBS and filtered through sterile 0.22-μm polyvinylidene difluoride filters (Millipore).

Cpl-1 and Pal (both WT and variants) for experiments including variants were supplemented with imidazole before incubation with NiNTA agarose up to 80mM. Then it was washed with 10x of 85mM of imidazole in PBS and active protein was eluted with 250mM of imidazole and dialyzed

against PBS at 4 °C (3,000 kDa molecular weight cut-off, 3 times, 2000 mL bottles for 20 mL of protein sample, 4°C, Spectra/Por, Repligen) against PBS. Samples were filtered through sterile 0.22-µm polyvinylidene difluoride filters (Millipore).

For experiments with mice, the final step was LPS removal, which was performed with an Endo-Trap Blue column (Hyglos GmbH). Purified, endotoxin-free protein samples were dialyzed against PBS and filtered through sterile 0.22-µm polyvinylidene difluoride filters (Millipore). The LPS (endotoxin) content was determined using the EndoLISA assay (Hyglos GmbH) and confirmed to be 0.6–1.6 endotoxin units (EU) per mL for all of the microarray gene expression profiling and less than 10 EU per animal in animal experiments.

All steps were monitored by SDS-PAGE at all stages. Protein concentration was determined by the Bradford assay (Sigma-Aldrich, Europe) following the manufacturer's instructions.

6. Microarray gene expression profiling

The cell lines FaDu (human pharynx squamous cell carcinoma, ATCC HTB-43) and SC (human peripheral blood macrophages, ATCC CRL-9855) were cultured in media with and without protein solutions containing Cpl-1 or Pal (0.5 µM, i.e. 17.5 and 20 µg/mL, respectively) for 6 h. This time was chosen to allow for the development of cellular responses to external factors at the cellular gene expression level. Control cells were cultured with an equivalent amount of albumin (Sigma). After incubation, cells were harvested and the total RNA was immediately isolated with RNeasy Mini Kit (Qiagen). For this study, we used SurePrint G3 Human Gene Expression v3.8 × 60 K Microarrays (Agilent Technologies). The One-Color Microarray-Based Gene Expression Analysis Protocol (version 6.9.1) was used to process the arrays.

After amplification, the total RNA was labeled with Cy3 using the Low Input Quick Amp Labeling Kit (Agilent Technologies, USA). The labeled RNA was purified (RNeasy Kit), and the RNA yield (nanograms of complementary RNA (cRNA)) as well as specific activity (picomoles of Cy3 per microgram of cRNA) were measured using a NanoQuant plate (Tecan Group) in an Infinite 200 PRO reader. Next, the labeled cRNA was fragmented and placed on the microarray slide after mixing with the hybridization buffer. Microarrays were incubated for 17 h at 65 °C and then washed twice in GE wash buffer (Agilent). Agilent's High-Resolution C Microarray Scanner was used to scan the slides according to the 8 × 60 K array format. The scanned images were analyzed with the Agilent Feature Extraction software v. 12.1 (Agilent Technologies). The final analysis included dye normalization (linear and LOWESS), background subtractions, and filtering of outlier spots.

7. Differentially expressed genes

GeneSpring GX 13 (Agilent) was used to further analyze the data after extraction. The cut-off was set to 1.5-fold for the determination of significant differential expression (up or down regulation). A moderated Student's *t* test was used to determine significant differences, defined as $p \leq 0.05$, for gene expression.

8. Enriched gene ontology terms & pathway analysis

Lists of differentially expressed genes were uploaded to the DAVID 6.7—Database for Annotation, Visualization and Integrated Discovery Classification System to analyze their ontologies and participation in curated pathways, and to perform a Gene Set Enrichment Analysis (GSEA). Ontologies and pathways were assigned independently to upregulated and downregulated gene lists. Annotations (official gene symbols) were limited to *Homo sapiens* and the genome of this species was used as a background for analyses. For the pathway analysis, a Kyoto Encyclopedia of Genes and Genomes (KEGG) was used as a reference database.

9. Animal experiments (safety experiments)

Six- to twelve-week-old male mice, either C57BL6/J (N = 6) or BALB/c (N = 5 or 7), were bred at the Animal Breeding Centre of the Institute of Immunology and Experimental Therapy (IIET). All the animal experiments were approved by the 1st Local Committee for Experiments with the Use of Laboratory Animals, Wrocław, Poland (projects no. 76/2011) and performed according to EU directive 2010/63/EU. All the animal experiments adhered to the ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines(59).

For the general health condition (section 2.5), the microbiome assessment (section 2.6), and the inflammatory/cytokine assessment (section 2.7), BALB/c mice were injected with Pal or Cpl-1 intraperitoneally in one dose at 0.3 mg per mouse (15 mg/kg) for each protein. The negative control was inoculated with PBS and a positive control of inflammation was inoculated with LPS (2000 EU/mouse). Murine blood was collected into clotting tubes under anesthesia from the tail vein. Serum was separated from the blood using double centrifugation (2250× g for 5 min and 10,000× g for 10 min). The sera were stored at −20 °C.

10. Overall health scoring matrix of mice

A scoring matrix indicating the general health of mice was calculated on a 15-point scale by adding the scores of five aspects, each rated from 0 to 3 (with mid-values possible), where the highest score represented the worst condition. The specific criteria used included: Activity (0 = alert and active, or calm and resting; 3 = slow to move or non-responsive when coaxed or violent reaction to stimuli); Fur (0 = normal, well groomed; 3 = very rough hair coat); Eyes (0 = normal, open, clean, no exudate; 3 = closed, sunken or covered with suppurate exudates); Abdomen (0 = normal; 3 = large abdominal mass and/or edema); and Skin (0 = normal, healthy skin; 3 = wounds, dermatitis, lesions).

11. Cytokine assay

The progression of the inflammatory reaction in the murine blood was monitored by measuring interleukin-6 (IL-6) serum levels using a commercially available Human/Mouse IL-6 Mini ABTS ELISA Development Kit (PeproTech Inc.) following the recommendations of the manufacturer.

12. Specific Sera Induction

For specific IgG induction, C57BL6/J mice were challenged with Pal or Cpl-1 (0.05 mg per mouse) or PlyC (0.1 mg/L) on day 0. Murine blood was collected into clotting tubes under anesthesia from the tail vein every 3–5 days. Serum was separated from the blood using double centrifugation (2250× *g* for 5 min and 10,000× *g* for 10 min). The sera were stored at −20 °C.

13. Measurement of specific IgG and IgE antibody levels

MaxiSorp flat-bottom 96-well plates (Nunc, Thermo Scientific) were sterilely coated overnight at 4 °C with endolysins or PBS as control using 100 µL per well at a concentration of 10 µg/mL. Subsequently, wells were washed 5 times with PBS and blocked for 1 h with 1% albumin or SuperBlock Blocking buffer (Life Technologies) at 150 µL per well at room temperature. Solution was removed and the plates were washed 5 times with 0.05% Tween 20 (AppliChem GmbH) in PBS at 100 µL per well. One hundred microliters per well of diluted serum (1:100 in PBS) was applied to the wells coated with endolysins as well as control wells. Each sample was investigated in duplicate. The plates were incubated at 37 °C for 2 h after which serum was removed and the plates were washed 5 times with 0.05% Tween 20 in PBS at 100 µL per well. One hundred microliters per well of diluted detection antibody (peroxidase-conjugated goat anti-mouse IgG (Jackson ImmunoResearch Laboratories) or IgE (ThermoFisher) was applied to the plates and incubated for 1 h at room temperature in the dark. The antibody solution was removed and the plates were washed with PBS with 0.05% Tween 20 5 times at 100 µL per well. TMB (50 µL) was used as a substrate rea-

gent for peroxidase according to the manufacturer's instructions (R&D Systems) and incubated for 30 min. Twenty-five microliters of 2N H_2SO_4 was added to stop the reaction and the absorbance was measured at 450 nm (main reading) and normalized by subtracting the background absorbance at 550 nm.

As a reference level of specific IgE antibody induction, an oral mouse allergy model to ovalbumin (OVA) was utilized. Allergy to OVA was induced in mice using a dedicated adjuvant as described(60). Briefly, the mice were injected subcutaneously with OVA (50 μg/mouse) with $Al(OH)_3$ as an adjuvant promoting the hypersensitivity reaction. The injection was repeated after 14 days. Seven days after the second sensitization, the mice were given 20% Egg White Solution (EWS) (Sigma, Poznan, Poland) in the drinking water. IgE levels specific for OVA were evaluated using ELISA as described above and compared to control mice injected with PBS instead of OVA.

14. Complement system activity test

Complement assays were performed with the WIESLAB® Complement System kits (Euro Diagnostica) to test the activation of the three complement pathways in the presence of Pal or Cpl-1. Briefly, the kits contained pre-coated plates with specific activators for the classic pathway (CP), alternative pathway (AP), or mannose-binding lectin (MBL) pathway (LP). Blood samples were collected from six healthy human donors. Blood was collected in clotting tubes and incubated for 60 min at room temperature, followed by 10 min centrifugation at 2000× *g*, and isolated serum was immediately used for the tests. First, human serum was incubated at 37 °C for 10 min with 2 μg/mL of Pal, Cpl-1, or PBS in equal volumes. Then, sera were diluted with the provided buffers at 1: 200, 1:18, 1:100 for the CP, AP and LP pathways, respectively, and transferred to the WIESLAB plate and processed according the manufacturer's guidelines. Positive and negative controls included in the test kit served as quality control of the assay.

15. Testing activity, fluorometric assay

Sytox Green solution from ViaGram™ Red+ Bacterial Gram Stain and Viability Kit (ThermoFisher Scientific) was used in this article. Kinetic experiments were performed at room temperature (23°C). Several controls were used to ensure technical suitability and reproducibility of experiments. Briefly (detailed description below), in each experiments we measured fluorescence of PBS (technical control), killed bacteria with isopropanol (positive control), live bacteria only (negative control), lytic agent only (enzyme control) and bacteria killed by endolysins (lysis control) showing the level of fluorescence when lytic reaction is complete.

The technical control (background) measures fluorescence of the background coming from the buffer, the plate and any unbound fluorescent dye. It consists of 150 µL PBS mixed with 50 µL of diluted (1/100) Sytox Green solution. The mean signal of this control was subtracted from all reads.

The positive control is used to demonstrate that the DNA dye in the sample is of sufficient quantity (e.g., not saturated) and the measurements are valid. This control is comprised of a concentrated suspension of dead bacteria that gives extremely high fluorescent reads. The signal measured by this control sample should be higher than in any other sample, establishing that the DNA dye is not saturated in the experimental groups and are technically valid. Preparation of the concentrated suspension of dead bacteria is based on a protocol provided by manufacturer of the DNA dyes (ThermoFisher). Briefly, 1 mL of S. pneumonia grown for 22 hr at 37°C in Todd-Hewitt broth without shaking was added to 20 mL of 70% isopropyl alcohol and incubated for 1 hr at room temperature with mixing every 15 min, followed by centrifugation (15,000 x g, 15 min, 4°C) and resuspension in 0.25 mL of PBS. 50 µL of this suspension was mixed with 50 µL of diluted (1/100) Sytox Green and 100 µL of PBS.

The negative control (bacteria) included 100 µL PBS with 50 µL of diluted (1/100) Sytox Green and 50 µL of previously prepared bacteria (OD600=1.0). This control was used for normalization to calculate progress equal to 0.0 for each timepoint.

The enzyme control consists of 100 µL PBS with 50 µL of diluted (1/100) Sytox Green and 50 µL of previously prepared enzyme solution in PBS (at the highest concentration used for the experiment).

The lysis control (for calculation of max signal in normalization) contains 50 µL PBS, 50 µL of diluted (1/100) Sytox Green, 50 µL of previously prepared enzyme (80 mg/L for Cpl-1 and 40 mg/L for Pal enzyme) and 50 µL of bacteria solution. The max value is defined as the mean signal measured in the last 3 minutes of the reaction. If lysis has gone to completion, the progress curves should show no statistically significant change in the last 3 minutes (i.e., the curve is flat). Enzyme concentrations included in the experiments higher than proposed can be used for calculation of the max signal. The amount of lysin used for this control was chosen such that complete lysis (99.9%) occurred within 10 min. Since activity of the bacteriolytic agent depends on bacteria strain, amount of the bacteriolytic agents needs to be sufficient for complete lysis in 10 min. For experiments presented in this article, the lysis of more than 99.9% of bacteria was confirmed by dilution testing.

1. Before experiment:

1.1 Each sample is tested in triplicate.

1.2 Prepare concentrations of the enzymes remembering that for a typical assay the enzyme is twice diluted after being added to the sample, so stock solutions of the enzyme needs to be twice the final desired concentration.

2. Bacteria preparation:

2.1 Susceptible bacteria of S. pneumoniae were grown for 20 h at 36◦C in (50 mL) in THY broth, no shaking.

2.2 Bacteria were harvested by centrifugation (7 000 x g, 7 min, 20°C).

2.3 Supernatant was removed.

2.4 Pellet was resuspended gently in PBS (25 mL).

2.5 Sample was centrifuged again (7000 x g, 7 min, 20°C)

2.6 Pellet was resuspended gently in PBS (25 mL).

2.7 Sample was centrifuged again (7000 x g, 7 min, 20°C)

2.8 Pellet was resuspended in PBS to final OD_{600} 1.0.

2.9 The sample was incubated for 30 min at a room temperature.

2.10 The sample is now ready for fluorescent measurements (point no. 4).

3. Controls preparation (each should be in triplicate per plate).

3.1 Positive control (should be started 2 h before the experiment):

3.1.1 Sytox Green solution from ViaGram™ Red + Bacterial Gram Stain and Viability Kit (ThermoFisher, alternatively - 5 mM solution of Sytox Green in DMSO can be used) was diluted in PBS 100 times (volume needed is 50 µL per well).

3.1.2 10 mL of bacteria from step 2.10 was centrifuged (7 000 x g, 4°C, 7min).

3.1.3 The supernatant was discarded.

3.1.4 The pellet was resuspended in 1 mL of PBS.

3.1.5 20 mL of 70% isopropyl alcohol was added (water solution).

3.1.6 The sample was incubated 1 h at room temperature and mixed every 15 min.

3.1.7 The sample was centrifuged (15 000 x g, 15 min, 4◦C).

3.1.8 The supernatant was discarded.

3.1.9 The pellet was resuspended in 350 µL of PBS.

3.1.10 50 µL of PBS with 50 µL of Sytox Green solution from step 3.1.1 and 100 µL of sample from point 3.1.9 was mixed in the well before the experiment.

3.2 *Technical control*: 150 µL of PBS and 50 µL of Sytox Green solution from step 3.1.1 was mixed.

3.3 *Lysis control*: 50 µL of PBS, 50 µL of Sytox Green solution from step 3.1.1, 50 µL of bacteria from step 2.10 and 50 µL of endolysin (at least 80 µg/mL for Cpl-1 and 40 µg/mL for Pal) was mixed.

3.4 Enzyme background control: 100 µL of PBS with 50 µL of Sytox Green solution from step 3.1.1 and 50 µL of endolysin (equal to highest concentration used in the experiment) was mixed.

3.5 Negative control (bacteria without endolysin): 100 µL of PBS, 50 µL of Sytox Green solution from step 3.1.1, 50 µL of bacteria from step 2.10 was mixed.

4. Fluorometric measurement.

4.1 To each well, 50 µL of Sytox Green solution from step 1, and 100 µL of endolysin stock solution in PBS (two times higher than tested) was added.

4.2 All necessary controls were prepared and mixed.

4.3 Enzymatic lysis was started by adding 50 µL of bacteria suspension from step 2.10 to each well from step 4.1.

4.4 Read was started immidately (ex./em.: 475/525 nm) each well every 45-60 sec.

4.5 Measurements continued for at least 10 min.

Validity of the experiment - results were considered valid only if:

1. Average raw signal of technical control is lower than the average signal enzyme control.

2. Average raw signal of the enzyme control is lower than average signal of negative control.
3. Average raw signal of positive control yields at least 33% higher reads than any other sample in the experimental setup.
4. Average raw signal from lysis control from last 3 minutes of the experiment shows no statistically significant change after the initial rise.

16. Testing activity, turbidity reduction assay (TRA)

As a comparison for the fluorometric assay, turbidity reduction assays were performed in parallel. The same protocol was followed with only the diluent (DMSO, 1% dilution in PBS) used in place of the solution of fluorescent dye. Otherwise, the same freshly prepared bacterial cells were utilized as was the same concentration range of endolysins. Technical and negative controls were the same as in the fluorometric assays. Samples were tested in standard 96-well plates in triplicate and absorbance readings were measured every for 10 min at 600 nm.

17. Bacteria used for bacteriolytic experiments

Clinically isolated Streptococcus pneumoniae strain (code PN135) was used for all experiments. It is susceptible to optochin, Pal and Cpl-1.

18. Epitope analysis

Identification of amino acids interacting with specific IgGs was performed in accordance with the protocol published by Xu et al. adapted to our research by using coding sequences of the investigated endolysins as the source for library design(61). The peptide library was created by selecting 56 amino acid (aa) long fragments tailing through Pal and Cpl-1 endolysins, starting every 10 aa from the first amino acid (if necessary, oligopeptides with the end of the protein were shorter). In addition to the WT peptides, variants with consecutive single alanine substitutions, double alanine substitutions (i.e. targeted amino acid plus a neighbouring amino acid) and triple alanine substitutions

(targeted amino acid plus both neighbouring amino acids) were also created. Alanine residues in original sequences were substituted by glycine. Thus, our library contains all possible single-, double- and triple substitutions for each position in the sequence for all oligopeptides designed as overlapping parts of the investigated endolysins.

The peptide library was reverse translated into DNA sequences using codons optimized for expression in *E. coli*. The oligonucleotide library was synthesized using the SurePrint technology for nucleotide printing (Agilent). These oligonucleotides were used to create a phage library representing all oligopeptides using the T7Select 415-1 Cloning Kit (Merck Millipore).

Immunoprecipitation of the library was performed in accordance with a previously published protocol.11 Briefly, the phage library was amplified in a standard culture, purified by PEG precipitation and dialyzed against Phage Extraction Buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 6 mM $MgSO_4$). All plastic equipment used for immunoprecipitation including Eppendorf tubes and 96-well plates were prepared by blocking with 3% BSA in TBST buffer overnight on a rotator (400 rpm). A sample representing an average of 10^5 copies of each clone was mixed with 2 µL of serum with high levels of specific anti-endolysin IgGs. This mix was prepared in 250 µL phage extraction buffer and incubated overnight at 4°C on a rotator (400 rpm). A 20 µl aliquot of a 1:1 mixture of Protein A and Protein G Dynabeads (Invitrogen) was added and incubated for an additional 4 h at 4°C on a rotator (400 rpm). Liquid in all wells was separated from Dynabeads on a magnetic stand and removed. Beads were washed 3 times with 150 µL of a wash buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% Tween-20) and beads were resuspended in 60 µL of water.

The immunoprecipitated part of the library was then used for amplification of the insert region (according to the manufacturer's instructions) with a Phusion Blood Direct PCR Kit (Thermo Fisher

Scientific). The primers T7_Endo_Lib_LONG_FOR GCCCTCTGTGTGAATTCT and T7_Endo_Lib_LONG_REV GTCACCGACACAAGCTTA were used and a second round of PCR was carried out with the IDT for Illumina UD indexes (Illumina Corp.) to add adapter tags. Sequencing of the amplicons in accordance with Illumina next generation sequencing (NGS) technology was outsourced (Genomed).

19. Sequencing data analysis

Sequenced amplicons were mapped to original nucleotide library sequences by the bowtie2 software as published by Xu et al(61).

NGS sequencing reads were mapped by the bowtie2 software package with use of the originally synthesised nucleotide library as indexes (local mode)(62). A number of hits that mapped to each reference sequence (highest score) was counted (count, c).

The signal in each sample was calculated according to a formula (1).

$$\boldsymbol{s}_{ijm} = \frac{c_{ijm}}{\sum_{i \in I} c_{ijm}} \quad (1)$$

s – signal of i-th sequence in j-th serum sample and m-th technical replicate

c – number of reads mapped to *i*-th sequence in *m*-th technical replicate of *j*-th serum sample

I – set of all reference sequences (used as indexes to in mapping by the bowtie2 software)

The vast majority of clones were washed away and were not enriched, if detected at all thus, a zero-inflated negative binomial model was used to evaluate the probability of each signal to be random. Next, all p-values of oligopeptides detected in the sample were adjusted for multiple hypotheses (FDR method) (25). Relative signal (enrichment): an average signal in technical replicates of the sample divided by average signal in 'input samples' of the same sequence resulted in signal ratio (relative signal). A series of t-tests evaluated the significance of differences in signals for each oligopeptide separately between input samples (one group) and the tested sera.

20. Variant design

Every detected immunogenic amino acid was substituted in silico with every other possible amino acid and the difference in free folding energy (ddG) was calculated using the FoldX software(63). A three-dimensional model of Cpl-1 endolysin was downloaded from the PDB (Crystal structure of the modular the Cpl-1 endolysin complexed with a peptidoglycan analogue (wild-type endolysin)(64). The three-dimensional structure of Pal was not available in the databases, so it was created using the I-Tasser software(65). Variants were chosen based on the following assumptions:

a) amino acids that differed from wild-type in their electrostatic charge and chemical structure were selected to disrupt epitope interaction with specific antibodies;

b) the lowest energy state of the folded variant was preferred;

c) smaller amino acids were preferred for the substitution (due to lower steric tensions).

For design of point mutants based on this analysis, amino acid sequences were reverse translated using frequent *E. coli* codons. Gene sequences coding the variants were synthesized de novo and cloned into pBAD vectors (BioCat GmbH, Germany).

21. Measurement of specific IgG antibody levels (recognition of variants by wild-type specific antibodies)

MaxiSorp flat-bottom 96-well plates (Nunc, Thermo Scientific) were coated in sterile conditions overnight at 4°C with endolysins or PBS as a control using 100 µL per well at a concentration of 10 µg/mL or 80 µg/mL. Subsequently, wells were washed 5 times with PBS and blocked for 0.5 h with SuperBlock Blocking buffer (Life Technologies Europe BV, Bleiswijk, Netherlands) at 150 µL per well at room temperature. The solution was removed and the plates were washed 5 times with 0.05% Tween 20 (AppliChem GmbH, Darmstadt, Germany) in PBS at 100 µL per well. One hundred microliters per well of diluted serum (1:350 in PBS) was applied to the wells coated with endolysins. Each sample was investigated in five repeats. The plates were incubated at 37°C for 2 h, after which serum was removed and the plates were washed 5 times with 0.05% Tween 20 in PBS at 100 µL per well. One hundred microliters per well of diluted detection antibody (peroxidase-conjugated goat anti-mouse IgG (Jackson ImmunoResearch Laboratories, Cambridgeshire, UK) was applied to the plates and incubated for 1 h at room temperature in the dark. The antibody solution was removed and the plates were washed with PBS with 0.05% Tween 20 five times at 100 µL per well. TMB (50 µL) was used as a substrate reagent for peroxidase according to the manufacturer's instructions (R&D Systems, Minneapolis, MN, USA) and incubated for 30 min in the dark. Twenty-five microliters of 2N H_2SO_4 was added to stop the reaction and the absorbance was measured at 450 nm (main reading) and normalized by subtracting the background absorbance at 550 nm.

22. Ethical statement

All human subjects gave informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Bioethics Committee of the Regional Specialist Hospital in Wrocław (Project identification code: KB/nr 2/2017).

All animal experiments were performed according to EU directive 2010/63/EU for animal experiments and were approved by the 1st Local Committee for Experiments with the Use of Laboratory Animals, Wrocław, Poland (project no. 76/2011). The authors followed the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines(59).

9. Results

1. Development of a new method for measurement of bacterial lysis dynamics

Studies of endolysin activity are limited by traditional, however not enough robust nor informative techniques for measurements of bacteriolytic activity. Two most widely used methods are the dilution method and the indirect turbidity reduction assay (TRA). They both are not sufficient for precise measurement in time nor for comparisons of lytic activity between different variants and wild-type endolysins. Dilution method delivers information on a number of live bacteria in a specified time and it is not appropriate for precise detection of enzyme kinetics. Turbidity method requires relatively large amounts of bacteria and it is characterized with low precision. One of the major parameters that characterize lysis is the *lytic activity* that refers to the initial velocity of lysis in terms of enzyme kinetics and is calculated by dedicated mathematical models. If a lytic component with known activity is mixed with an unknown amount of bacteria, the lytic activity value can be referred to as the *susceptibility to lysis*. This setup can be used for characterization of bacteria mixtures and analysis of unknown samples, such as clinical or environmental isolates, however calculation of these parameters require a detection method that is enough sensitive, precise, and controllable to yield with precise mathematical values. For this reason I developed and validated new a lysis detection method that employs DNA-binding fluorescent dyes.

1. Sytox Green demonstrates the best applicability for detection of bacterial lysis when compared to DAPI, SYTO9 and propidium iodide

Potential dyes for bacterial lysis detection were chosen and their applicability to measurement of bacterial lysis were evaluated by comparison with TRA method. I selected the DNA dye suitable for the fluorometric assay with following assumptions:

1) A suitable DNA dye should provide a stable, linear signal for the negative control (bacterial sample, see Materials and Methods for details) as determined by a linear regression model. If the signal from the negative control (bacteria, without endolysin) changes over time, a linear regression model evaluates a coefficient of determination (R^2 value) and a statistical significance of the model showing probability that a measured change in time (slope) is accidental. A coefficient of determination (R^2), specifies the goodness-of-fit, that is it describes how much the developed model varies from the experimental data. The statistical significance of the model is expressed as a p-value, where $p<0.05$ means that the slope deviates statistically significantly from zero.

2) The DNA dye should enable a detection of a fluorescent signal reaching a plateau after the lysis reaction ends during the reaction time. This plateau was used later as the *max* signal (progress equal to 1.0, meaning lysis of all the cells in the sample). The same plateau is seen in the turbidity reduction assay. The DNA dye showing practically no change in the last 3 min of the measurement was deemed suitable for our purposes.

3) A suitable DNA dye should show high responsiveness, meaning it should rapidly enter dead cells and bind the bacterial DNA immediately, preferentially not requiring complete lysis of the cells by the bacteriolytic agent. The DNA dye showing the most rapid change of the fluorescent signal in response to a bacteriolytic agent added to a sample may provide the highest speed and precision of measurements. This comparison was evaluated by a minimal time required to detect the change of fluorescent signal equal to 50% of the total detected change.

Four commercially available fluorescent DNA dyes were tested, Sytox Green and DAPI from the ViaGram™ Red+ bacterial Gram Stain and Viability Kit, and SYTO9 and propidium iodide from the LIVE/DEAD® BacLight™ Bacterial Viability Kit. DAPI and SYTO9 DNA dyes are used

commonly to stain live cells. Sytox Green and propidium iodide are used to stain metabolically inactive cells, so a rise of the signal after addition of the endolysin is expected due to the increase in metabolically inactive cells.

Comparison of propidium iodide, SYTO9, Sytox Green and DAPI proved that Sytox Green was the most suitable for my purpose of bacterial lysis detection. For the other dyes, the coefficient of determination was higher (R^2=0.047 for Sytox Green, R^2>0.3 for other dyes) and the statistical significance of the slope being accidental was the highest (slope = 0; p=0.21 for Sytox Green, p<0.001 for other dyes). The plateau (*max* signal) was detected only for DAPI and Sytox Green. DAPI required almost 6 minutes to reach 50% of the measured reaction progress and Sytox Green required only 4 minutes, making this dye the best choice for further experimentation.

Raw data from fluorometric assay and TRA can be normalized to an objective value: *progress* [of the lysis]. This is required for comparison between experiments using different measurement methods, enzymes, concentration of bacteria and/or enzymes. The *progress* (x) is defined as a part of the bacterial sample that is lysed, with values ranging between 0 (no lysis) to 1 (complete lysis of 100%bacterial cells). *Progress* is calculated by equation (1). Lysis control was used to calculate max signal for the purpose of normalization.

$$x(t) = \frac{|r(t) - bc(t)|}{|max - bc(t_{max})|}$$

x(t)– progress at time t

r(t)– signal of the sample at time t

bc(t) – signal of bacterial control at time t

(t_{max}) –signal of negative control at 10 min (end of the experiment)

max – signal of lysis control (enzymatically lysed bacteria)

Progress values can be used to calculate *lytic activity* as a part of the total bacterial sample that is lysed per unit of time [min^{-1}]. Raw data representative for kinetic measurements from TRA and fluorometric assay are presented for endolysins Pal and Cpl-1 (Figure 1A). Results of normalization in accordance to equation no. 1 with use of the same data are show *progress* of lysis 0 what part of a bacteria sample is lysed at the specific time (Figure 1B).

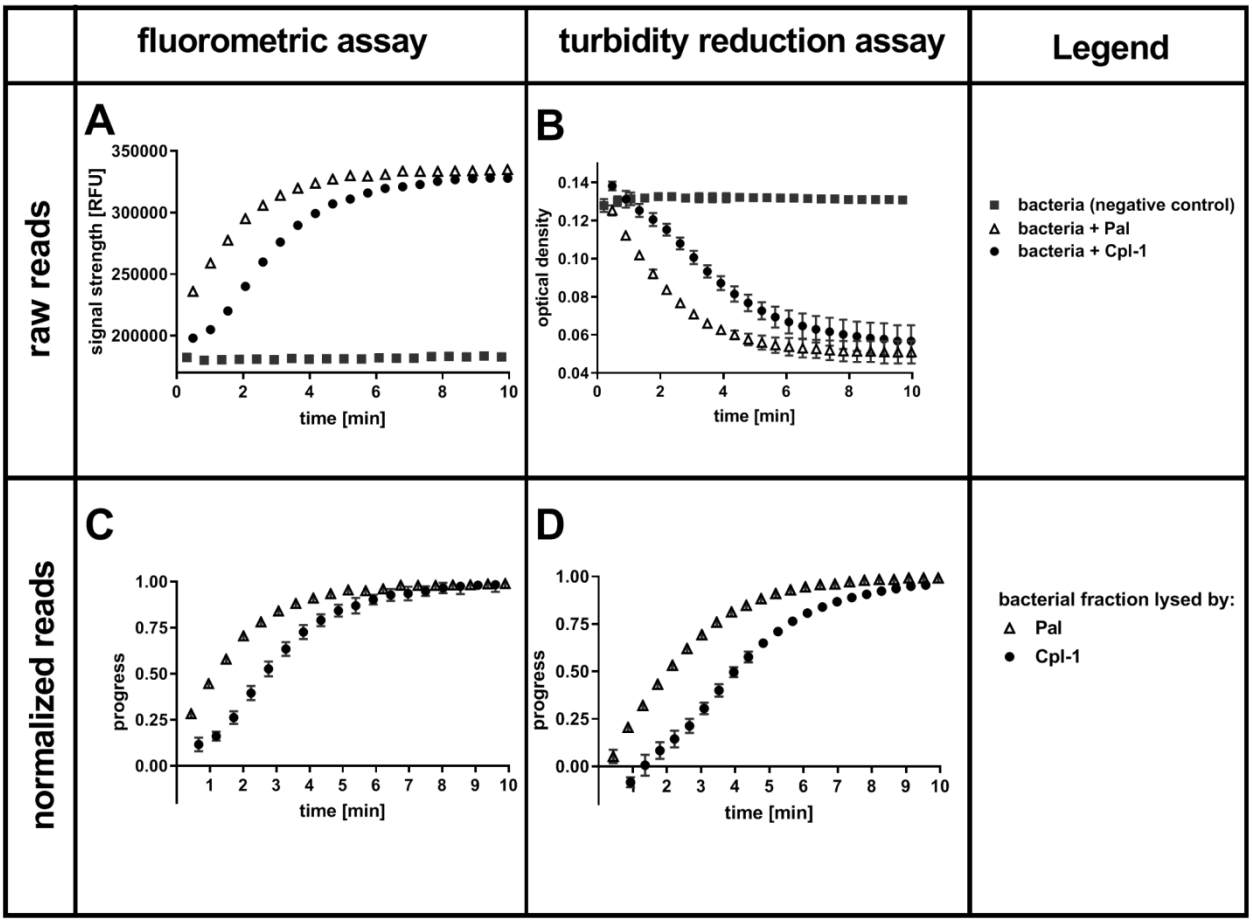


Figure 1. Fluorometric assay and turbidity reduction assay in a detection of bacterial lysis: representative raw measurements and normalized reads. Bacterial lysis was detected in PBS (200 µl) with 1.6x107 CFU, Cpl-1 or Pal (final concentrations of 46 mg/L and 13 mg/L, respectively) Sytox Green solution in DMSO or DMSO (fluorometric assay respectively or turbidity reduction). Reads measured in fluorometric (**A**) and turbidity assays (**B**). Data normalization: the bacterial fraction lysed (progress) calculated from reads for fluorometric (**C**) and turbidity reduction assays (**D**) (see: Equation 1; progress equal 0 means no lysis and progress equal 1 means all bacteria in a sample are lysed, for details see Materials and Methods). Average values from three technical repeats are demonstrated, whiskers represent the standard deviation.

The Cpl-1 lysin demonstrated a *lytic activity* that fulfils the requirements of a linear regression model. This model was calculated from *progress* values 0.2 to 0.75. In contrast, a one-phase association model best fit the data for the Pal endolysin (GraphPad Prism 7; R^2 >0,99 for both measurement methods) from 1 mg/L to 110 mg/L. *Lytic activity* in this model is evaluated by a K value, which depicts the change of *progress* at the beginning of the reaction (i.e. the initial velocity). In low concentrations of endolysins (below 1 mg/L of Pal and below 2 mg/L for Cpl-1), all measured points were used for calculation of linear regression due to insufficient lysis by the end of the 10 min assay. In these cases, the measured slope was used to define the *lytic activity* (Figure 2).

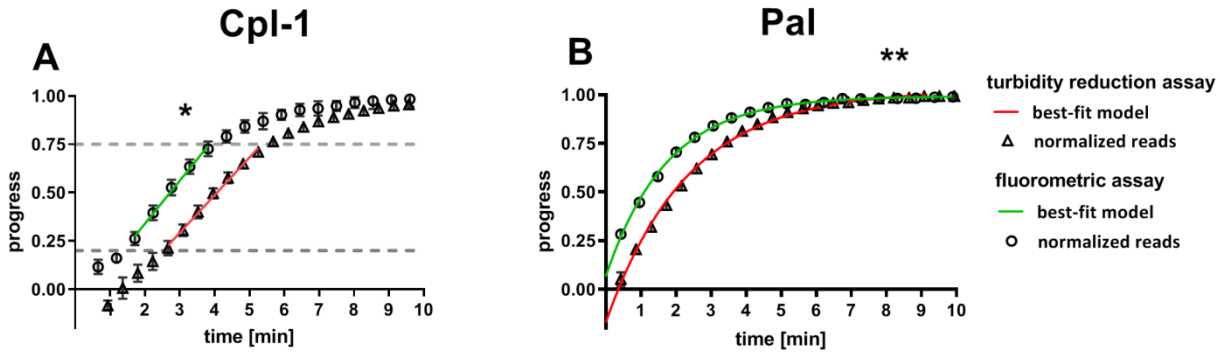


Figure 2. Fitted mathematical models for Cpl-1 and Pal lytic curves for exemplary data. Bacterial lysis was detected in PBS (200 µL) with $1.6x10^7$ CFU, Cpl-1 or Pal (final concentrations of 46 mg/L and 13 mg/L, respectively) and 0.25% DMSO or Sytox Green solution (turbidity reduction or fluorometric assay respectively). Graphical representation of mathematical models calculated from data presented in Figure 4 for Cpl-1 (**A**) and Pal (**B**). Proposed best-fit models allow for highly accurate (R2>0.99 in each model) and statistically significant (probability of inadequate model, p<0.0002, GraphPad Prism). Average values from three technical repeats are demonstrated and whiskers represent the standard deviation. Points between grey lines were used for the linear regression model of the Cpl-1 endolysin. All points measured for the Pal endolysin were used for mathematical modelling. * - the significant difference between lytic activity detected in fluorometric and turbidity reduction assays for Cpl-1 (p=0.048); ** - the significant difference between lytic activity detected in fluorometric and turbidity reduction assays for Pal (p<0.0001).

2. General mode of action for Sytox™ Green dye

Sytox Green fluorescent dye emits a signal after binding to double stranded DNA. This specific dye is also excluded from entering metabolically active bacteria (Figure 6). Taken together, observed fluorescence after addition of Sytox Green dye can arise from three sources: metabolically inactive (dead) bacteria, DNA contamination in the sample and background fluorescence. After addition of a bacteriolytic agent such as an endolysin, live bacteria are lysed (killed) and their DNA becomes available to Sytox Green. An increase in fluorescent signal that follows such lysis therefore directly corresponds to the activity of the bacteriolytic agent (Figure 3).

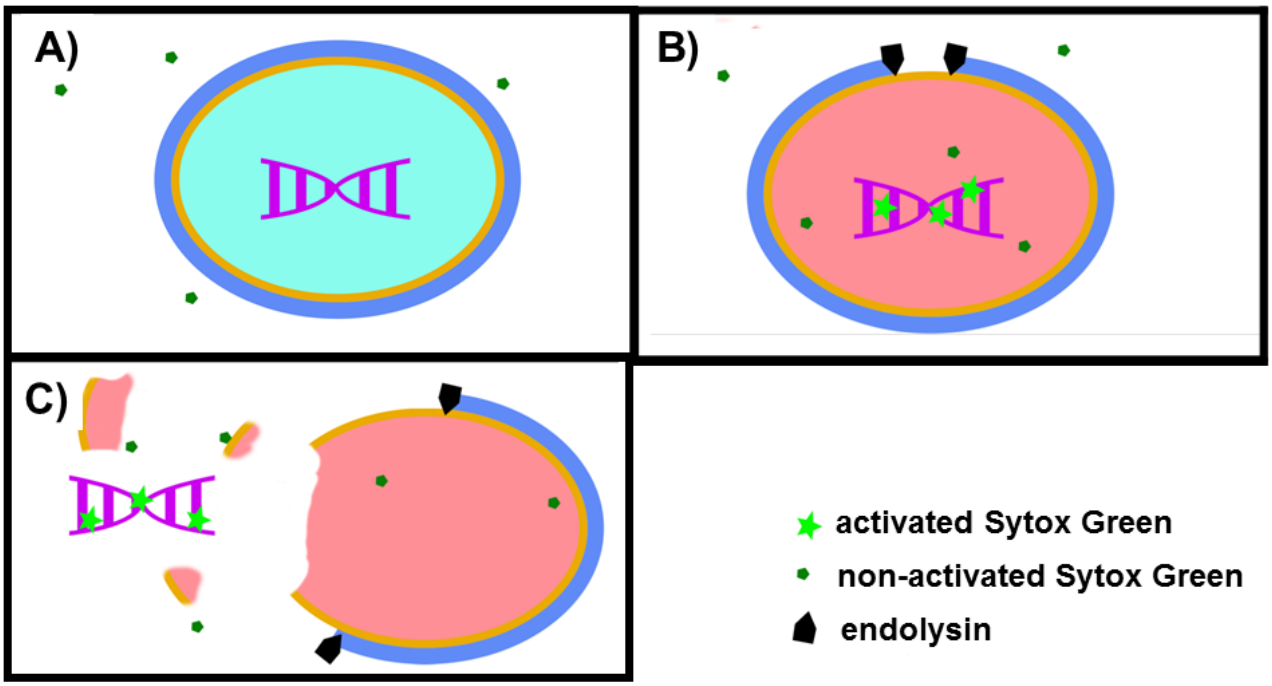


Figure 3: Mode of action for Sytox™ Green DNA dye. When added to a bacterial mixture, Sytox Green is excluded from metabolically active (live) bacteria and DNA remain sequestered inside the bacterial cell resulting in no signal (**A**). When a bacteriolytic agent (e.g., endolysin) is added to a mixture of live bacteria and Sytox Green, lysis of the bacterial cell begins. The bacterial cell becomes metabolically inactive and dies when sufficient lysis occurs, allowing Sytox Green to enter the cell ghost and bind to DNA resulting in increased fluorescence (**B**). In the case of extensive damage or even complete disintegration of the bacteria, bacterial DNA can leak outside the cell and Sytox Green in the mixture binds the DNA producing a signal (**C**).

3. Fluorometric assay measures a lytic activity of endolysins with a higher responsiveness and in a wider range of enzyme concentrations than the turbidity reduction assay

Lytic activities, *relative lytic activities*, and *specific lytic activities* were calculated from fluorometric assay and compared to turbidity reduction assays (TRA). Results are summarized in Figure 4. For the first time Cpl-1 and Pal endolysins were characterized with presented three measures:

A *lytic activity* is expressed as progress (part of the bacterial sample lysed per minute).

A *relative lytic activity* is a lytic activity presented as a percentage of lytic activity detected by the turbidity reduction assay under the same conditions.

A *specific lytic activity* represents activity of 1 nmol of the endolysin in a specific experimental set-up.

Lytic activity was tested in a range of enzyme concentrations, ranging from 0.18 to 375 mg/L for Cpl-1 and 0.05 to 110 mg/L for Pal. *Lytic activity* was detected by Sytox Green as low as 0.36 mg/L for Cpl-1 and 0.1 mg/L of Pal, while the minimum concentrations shown to have activity with the turbidity reduction assay were four times higher. These values were found significant in comparison to negative control (bacteria, no bacteriolytic agent) ($p < 0.001$ and $p < 0.008$, for Cpl-1 and Pal, respectively). The *lytic activity* detected by fluorometric assay was significantly higher than that detected by turbidity reduction assay, thus demonstrating higher responsiveness of the Sytox Green-based method. For Cpl-1, statistically significant differences between assays were observed only at concentrations 11.7 mg/L and lower ($p < 0.05$) (Figure 4A), whereas significance between assays was detected in all concentrations of Pal ($p < 0.05$) (Figure 4B).

A *relative lytic activity* allows for comparison of the two methods by artificially setting the activity detected by the turbidity reduction assay to 100% for a given set of conditions (Figures 4C,D). In doing so, differences in *lytic activity* at lower concentrations of a bacteriolytic agent are revealed. Such differences might be important in situations where only low concentrations of an enzyme are possible (e.g., due to side effects, cost, poor penetration to the target site). This analysis also revealed that *relative lytic activity* detected by fluorometric assay was significantly higher than that detected by turbidity reduction assay, demonstrating higher responsiveness of the Sytox Green-based method. Statistically significant differences between assays were observed for Cpl-1 only at concentrations 11.7 mg/L and lower ($p < 0.05$) (Figure 4C), while significant differences were detected for Pal in all concentrations ($p < 0.05$) (Figure 4D). Analysis of correlation *between relative lytic activity* and enzyme concentrations for Cpl-1 revealed negative correlation ($p < 0.0001$, Spearman correlation). However, the correlation for the Pal endolysin was not observed ($p > 0.18$, Spearman correlation), therefore, we calculated average *relative lytic activity* of Pal, which was 180 ± 23% ($p < 0.005$).

A specific lytic activity, as revealed by fluorometric assay and turbidity reduction assay, differed markedly (Figures 4E,F). The fluorometric assay detected up to 2.44 times *higher specific lytic activity* for Cpl-1 and 3.19 times higher for Pal in comparison to the turbidity reduction assay depending on the enzyme concentration. A drop in *specific lytic activity* at low enzyme concentrations are related to detection limits for concentrations below 0.36 mg/L for Cpl-1 and 0.1 mg/L of Pal. Conversely, a drop in *specific lytic activity* at high enzyme concentrations (Figures 4E,F) likely relates to inhibitory effects due to many endolysin molecules binding surface of one bacteria despite only a single lytic event possible for each bacterial cell. I recommend evaluation of a max *specific lytic*

activity for an endolysin for purposes of comparison between enzymes (2.75 ± 0.09 for Cpl-1 WT, 5.7 ± 1.4 for Pal WT).

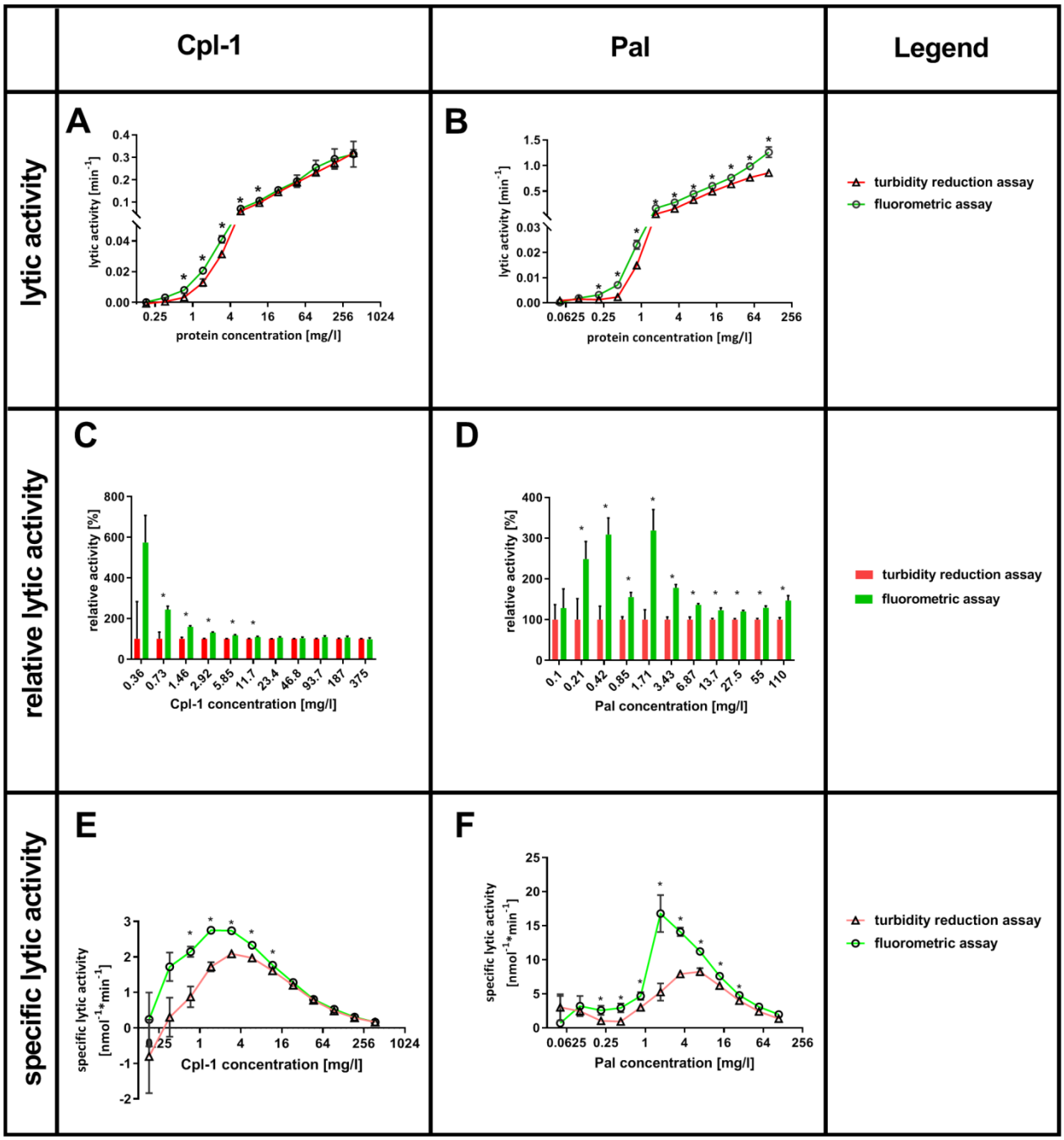


Figure 4. Comparison of Sytox™ Green fluorometric assay and turbidity reduction assay in characterization of Cpl-1 and Pal. Bacterial lysis was detected in PBS (200 mL) with 1.6×10^7 CFU, Cpl-1 or Pal at indicated concentrations and 0.25% DMSO or Sytox Green solution (turbidity reduction or fluorometric assay, respectively). Lytic activity of Cpl-1 (**A**) and Pal (**B**) is measured by both methods in concentrations ranging from 0.05 to 110 mg/L for Pal and from 0.18 to 375

mg/L for Cpl-1. Comparison of relative lytic activities for Cpl-1 (**C**) and Pal (**D**). Specific lytic activity of Cpl-1 (**E**) and Pal (**F**) measured by both methods. * - statistically significant difference in activity as detected by the two methods (adj. $p < 0.05$). Points represent results of two identical experiments. Bars represent the standard error.

4. Fluorometric assay detected bacterial lysis in a wider range of bacterial concentrations than the turbidity reduction assay (3.4×10^3 CFU/mL)

The lytic activity of Cpl-1 was detected in all measured bacterial concentrations, from as low as 3.4×10^3 CFU/mL (6.8×10^2 CFU/sample) to 8×10^7 CFU/mL (1.6×10^7 CFU/sample) in the fluorometric assay (Figures 5A,B). In contrast, optical density based methods, including the turbidity reduction assay, struggle to measure bacterial concentrations lower than 10^6 CFU/mL (66–69). Correlation between bacterial concentration and the detected signal of lysis was very high (0.9989, $p < 0.0001$ when 1.0 describes ideal correlation, Spearman correlation, GraphPad Prism 7) (Figures 5A,B). We assessed potential dependence of Cpl-1 lytic activity detected by fluorometric assay on bacterial concentrations and we found no statistically significant correlation ($p > 0.2$). Mean lytic activity was constant for all bacterial concentrations exceeding 5×10^3 CFU/mL, not deviating from 0.192 (95% CI from 0.1849 to 0.1985 min^{-1}; $p < 0.01$, GraphPad Prism 7). The fluorescent signal was measured after 10 min lysis by Cpl-1 over a range of bacterial concentration from 1.6×10^7 CFU/sample to 6.8×10^2 CFU/sample (Figure 5C). Bacterial lysis was detected in PBS (200 μL) with bacteria at indicated concentrations and 0.25% Sytox Green solution.

However, lytic activity of Cpl-1 in bacterial concentrations lower than 5×10^3 CFU/mL deviated from the used model for Cpl-1, but one-phase association model for this data was statistically significant in comparison to linear regression (adj. R^2 of 0.802 to 0.608 respectively). This effect likely resulted from a relatively large contribution of signal generated by the endolysin in comparison to other bacteria concentrations in which protein fluorescence is negligible (Figure 5C). This contribu-

tion is a background signal from protein added at the start of the lytic reaction. Most probable cause of this are some minor non-specific interactions between the protein and the dye. Nonetheless, this contribution is lower than $1/1000^{th}$ of the activated dye and is considerable only in this extremely low concentration of bacteria. Thus, this model was used for evaluation of *lytic activity* at the lower limit of bacterial concentrations (Figure 5C).

5. Optimal concentration of bacteria for fluorometric assay was between 2.5×10^7 and 2.5×10^6 CFU/mL per sample

The optimal concentration of substrate bacteria was determined experimentally to further improve experimental design for practical applications. Signal to noise ratio for different bacterial concentrations for the Cpl-1 endolysin was determined. An optimal range in the presented experimental setup was between 5×10^6 and 5×10^5 CFU, where the final signal reached a value over 12 times higher than the background signal value (Figure 5D).

6. The Cpl-1 endolysin is characterized by Michaelis-Menten model

A lytic activity of the Cpl-1 endolysin in varying concentrations of bacteria was used to calculate Michaelis-Menten statistics to develop a molecular characterization of the enzyme. I have demonstrated that Cpl-1 can be described by the Michaelis-Menten model when measured by the fluorescent method. V_{max} represents theoretical upper limit for activity of the enzyme in an abundance of substrate. For Cpl-1, this was $7.2 \pm 0.2 \times 10^6$ CFU/min. The K_m represents the concentration in which an enzyme demonstrates half of its maximum activity. For the Cpl-1 endolysin, it was $51 \pm 3\times10^6$ CFU/mL (Figure 5E)

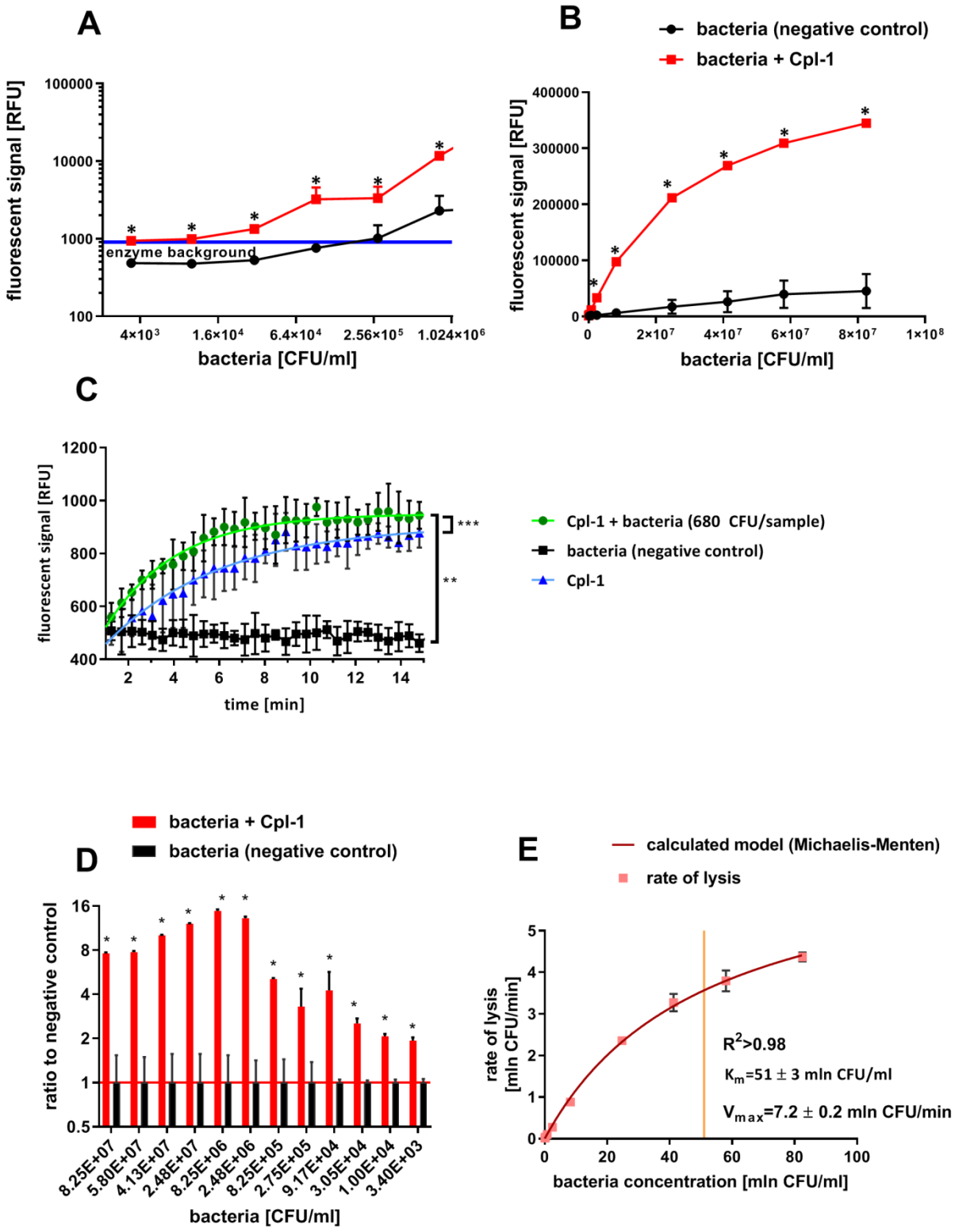

A
fluorescent signal [RFU]
100000
10000
1000
100
enzyme background
4×10³
1.6×10⁴
6.4×10⁴
2.56×10⁵
1.024×10⁶
bacteria [CFU/ml]
B
bacteria (negative control)
bacteria + Cpl-1
fluorescent signal [RFU]
400000
300000
200000
100000
0
2×10⁷
4×10⁷
6×10⁷
8×10⁷
1×10⁸
bacteria [CFU/ml]
C
fluorescent signal [RFU]
1200
1000
800
600
400
2
4
6
8
10
12
14
time [min]
***
**
Cpl-1 + bacteria (680 CFU/sample)
bacteria (negative control)
Cpl-1
D
bacteria + Cpl-1
bacteria (negative control)
ratio to negative control
16
8
4
2
1
0.5
8.25E+07
5.80E+07
4.13E+07
2.48E+07
8.25E+06
2.48E+06
8.25E+05
2.75E+05
9.17E+04
3.05E+04
1.00E+04
3.40E+03
bacteria [CFU/ml]
E
calculated model (Michaelis-Menten)
rate of lysis
rate of lysis [mln CFU/min]
5
4
3
2
1
0
20
40
60
80
100
bacteria concentration [mln CFU/ml]
R²>0.98
Km=51 ± 3 mln CFU/ml
Vmax=7.2 ± 0.2 mln CFU/min

Figure 5. Results of fluorometric assay in varying concentrations of bacteria and calculated Michalis-Menten model. Fluorescent signal of Sytox Green after 10 min of bacterial lysis in low (**A**) and high (**B**) concentrations of bacteria (red) and background signal of bacteria with no endolysin (black). Blue line shows background signal of Cpl-1 alone. Kinetics of Cpl-1 activity in the lowest concentration of bacteria (680 CFU/sample) in comparison to background signals from bacteria (negative control) and Cpl-1 only (blue line) (**C**). Ratio of signals from Cpl-1-lysed bacteria to negative control (bacteria) tested in a range of bacterial concentrations; the highest values represent optimal bacterial concentrations (**D**). Graphical representation of a Michaelis-Menten model calculated for the Cpl-1 endolysin with experimental data (**E**). * – statistically significant difference between samples and background ($p < 0.003$). ** and *** signifies statistically significant difference between K values of plotted curves (** is $p < 0.0001$, and *** is $p = 0.0269$). Points or bars represent average of measurements for two independent experiments and whiskers represent the standard deviation. A yellow line represent graphically K_m value.

2. Immunogenicity of endolysins Cpl-1, Pal, and PlyC in mouse model

Immune system is expected to produce specific antibodies as a response to exposure to proteins of non-human origin. Profile of antibodies induction by the investigated endolysins was assessed in mice by an intraperitoneal challenge with 50 μg/mouse (Cpl-1 and Pal) or 100 μg/mouse (PlyC) of a tested protein (one dose application). Higher amount of PlyC was applied due to its higher (than Cpl-1 and Pal) molecular mass. Murine sera were then collected for up to 50 days after exposure. Serum levels of specific antibodies were measured by ELISA test. Negative control consisted of a mouse group treated with PBS instead of endolysin solutions.

Immunogenicity was demonstrated in mice challenged with Cpl-1, Pal and PlyC; these animals showed a typical profile of specific IgG induction, that can be observed also for other proteinaceous antigens, including pathogen-derived proteins or vaccines. In comparison to negative control, these mice showed significantly higher levels of endolysin-specific IgGs on day 15 after administration of

Pal or PlyC, and on day 30 for Cpl-1 (adj. p< 0.05). After 30th day levels of IgG specific to the tested proteins reached a plateau (Figure 6A).

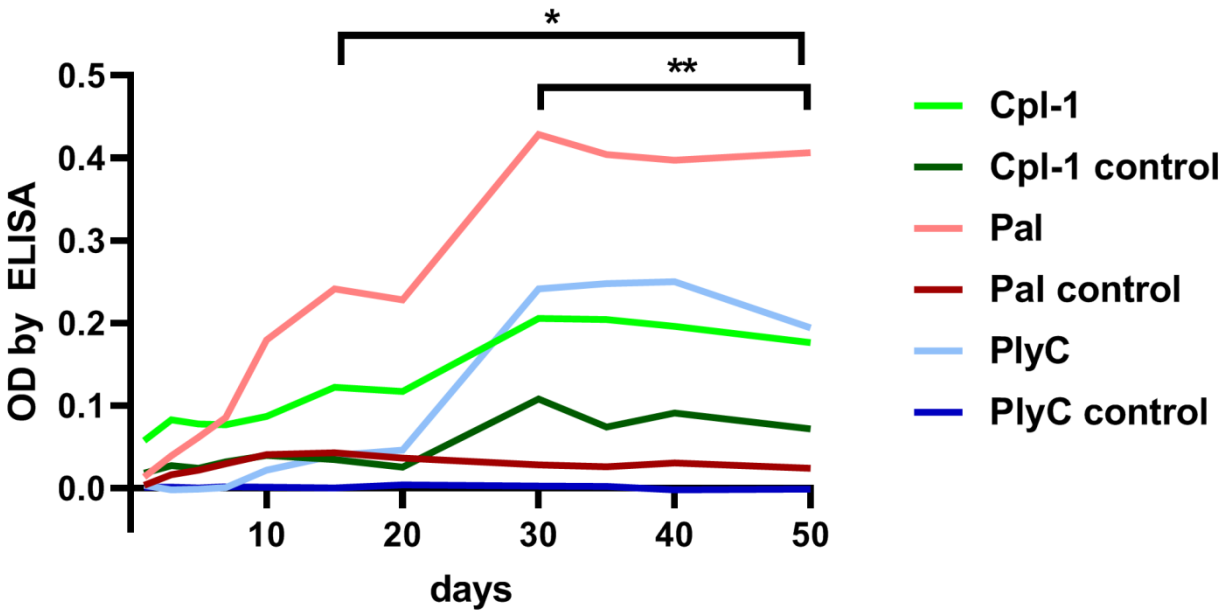


Figure 6. Antibody levels of IgG in mice challenged with Cpl-1, Pal and PlyC. Animals were exposed to purified protein by injection (i.p.). Serum samples were collected on days 1, 5, 10, 15, 20, 25, 30, and 50 and tested for levels of specific to Cpl-1, Pal or PlyC antibodies by ELISA. * - significantly different levels between groups exposed to PlyC or Pal tested group and respective control groups on the same day (adj. p < 0.05). ** - significantly different levels between a groups exposed to Cpl-1 and a respective controls on the same day (adj. p < 0.05). Each line represent average of antibody levels in a group (n=6 in each group).

3. Identification of immunogenic regions in endolysins - EndoScan technology

Identification of antigenic epitopes was done by epitope scanning (EndoScan, Figure 7), a new approach proposed herein for protein epitope identification, modified from the innovative VirScan technology developed by Xu et al.(61). Detailed specification of our EndoScan method is presented in Figure 3. Briefly, a T7 phage display library of 56 aa fragments of endolysins starting each subsequent 10 aa (46 aa overlap) was constructed. Murine sera were developed by challenging mice with the investigated endolysins and their respective sera with high titers of specific antibodies were collected. T7's library was then incubated with murine sera containing high levels of anti-Cpl-1, anti-Pal and anti-PlyC IgG. Complexes of specific IgGs attached to oligopeptides on the surface of T7 phages were then precipitated with magnetic beads binding to IgG (immunoprecipitation). Thus, a fraction of the phage display library that presented endolysin epitopes reactive to IgG present in the sera was isolated, while the rest, not recognized by IgG, was washed away. Since the library contains whole phage particles, each phage that displays a specific epitope contains a sequence coding for this epitope in its genome. Thus, NGS sequencing of the relevant region in the library genomes was used to reveal which epitopes were effectively recognized by specific IgG. Occurrences of each oligopeptide after immunoprecipitation were counted, normalized and compared to input library. Statistical significance was calculated between a group of sera from mice exposed previously to tested protein and input library.

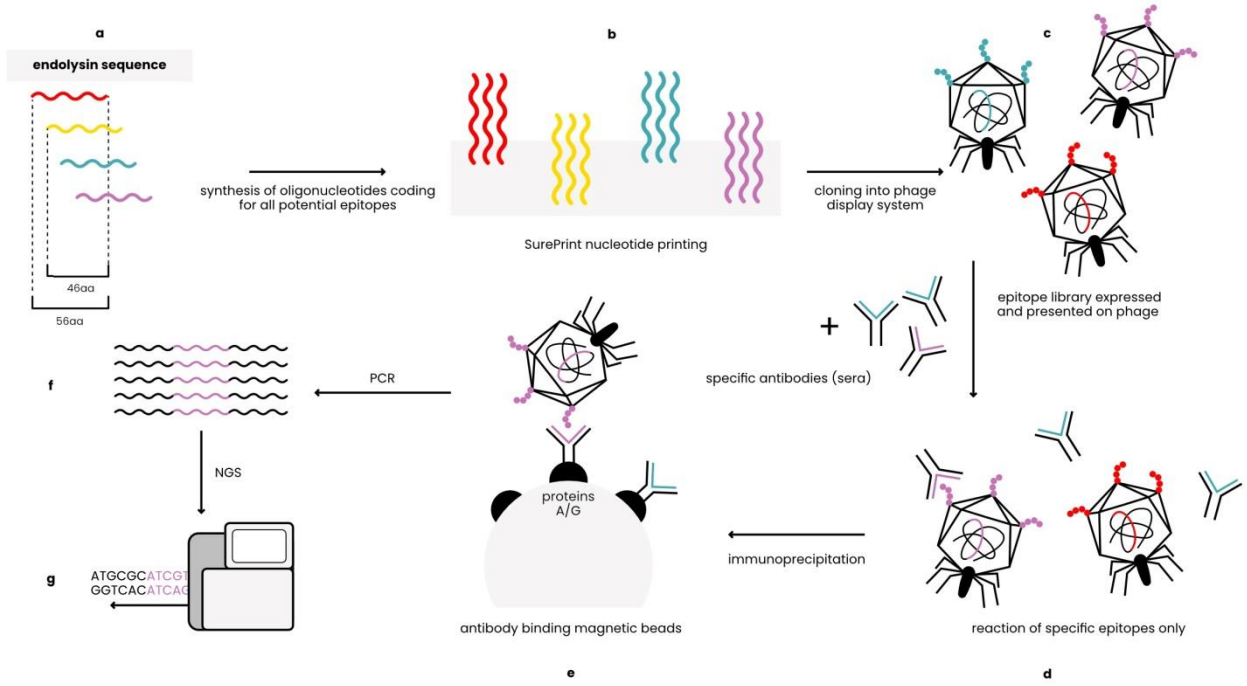


Figure 7. EndoScan technology. (**a**) *in silico* design of peptides covering sequences of Pal and Cpl-1, (**b**). synthesis of oligonucleotides coding for the peptides, (**c**). constructing of phage display library of endolysin-derived peptides, (**d**) reaction of the library with specific sera, (**e**) immunoprecipitation with magnetic beads binding Fc fragments of antibodies, (**f**) amplification by PCR reaction, (**g**) NGS sequencing, modified from Xu et al. (61).

1. Immunogenic regions of Pal, Cpl-1, and PlyC identified in a mouse model

Significantly increased reactivity with Cpl-1, Pal, and PlyCA specific IgGs was detected in 7, 3, and 7 oligonucleotides (within the EndoScan library) respectively, but no increased reactivity was detected in PlyCB subunit (Figure 8, adj. $p < 0.05$).

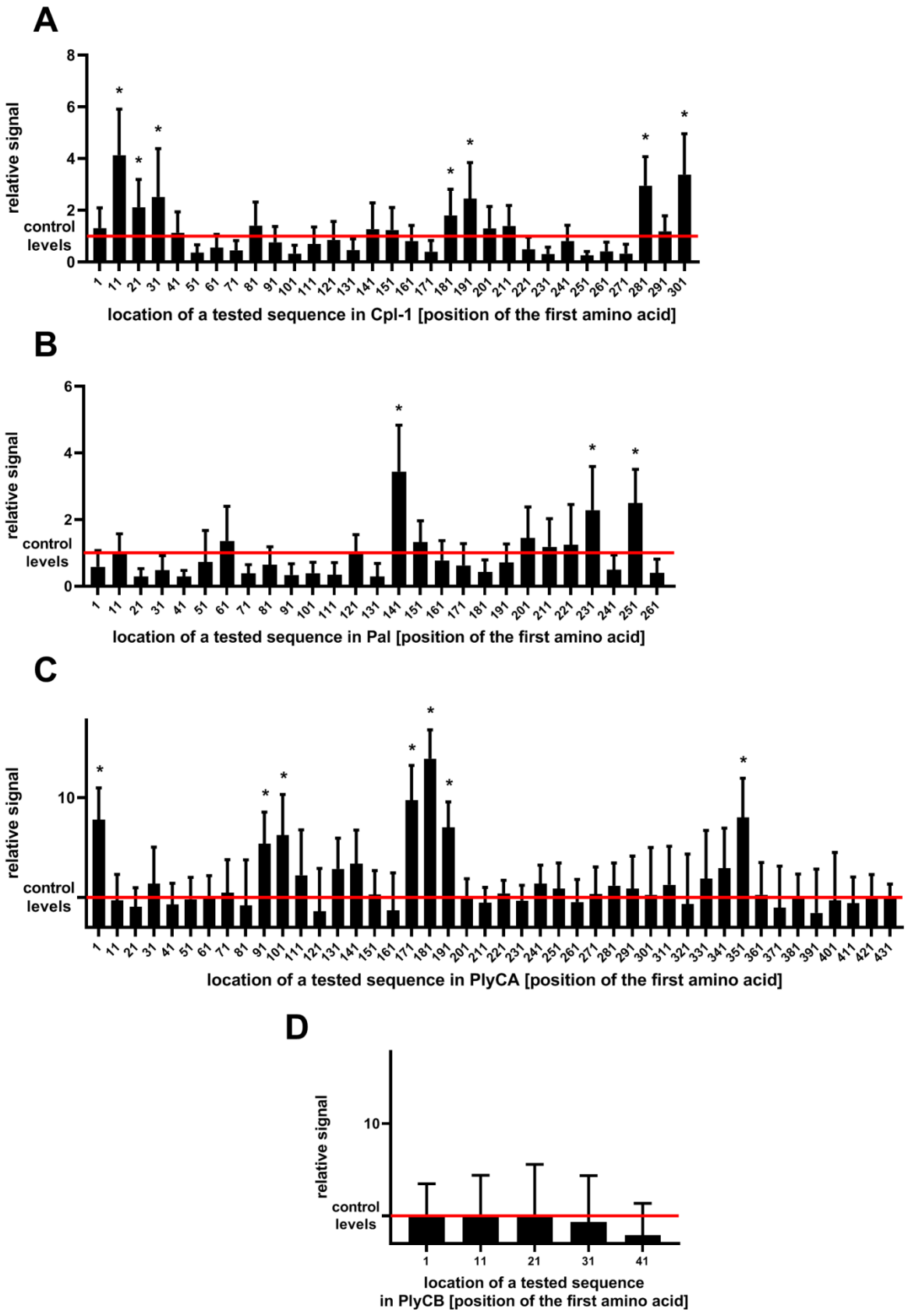

A
relative signal
control levels
location of a tested sequence in Cpl-1 [position of the first amino acid]
B
relative signal
control levels
location of a tested sequence in Pal [position of the first amino acid]
C
relative signal
control levels
location of a tested sequence in PlyCA [position of the first amino acid]
D
relative signal
control levels
location of a tested sequence
in PlyCB [position of the first amino acid]

Figure 8: Immunogenic regions in endolysins Cpl-1 (A) Pal (B), PlyCA (C) and PlyCB (D) identified with murine sera. Reactivity of murine endolysin-specific sera with phage-display library containing oligopeptides coding fragments tailing endolysin sequence (10 aa shift) has been presented. Endolysin-specific serum was developed in mice and was used for immunoprecipitation. NGS sequencing was used to identify the immunoprecipitated fraction of the phage display library. The relative signal of oligopeptides after immunoprecipitation is presented, indicating the relative strength of interaction between 56 aa oligopeptide and specific IgG. Red lines represent the signal of the 'input sample' (that is: the library before immunoprecipitation), normalized as 1.0 * - adj. $p<0.05$ from comparison between normalized signals from immunoprecipitated sample and input sample, for each oligopeptide. Datapoints represent average and SD of 6 biological replicates (n=6), two technical replicates were completed.

2. Immunogenic regions of Pal, Cpl-1, and PlyC identified in human sera

Phage library was also immunoprecipitated with human sera (n=56). This allowed to identify regions in Cpl-1, Pal and PlyC (Figure 9) that can be recognized by antibodies in humans non-exposed to endolysin. Such interactions may either represent unintended exposure for instance from natural components of human microbiome, or cross-reactions (false-positive recognition) by some IgG fractions induced by different antigens. Significantly increased reactivity with Cpl-1, Pal, and PlyCB specific IgGs was detected in 8, 4, and 2 oligonucleotides respectively, but no increased reactivity was detected in PlyCA subunit (Figure 3, adj. $p <0.05$).

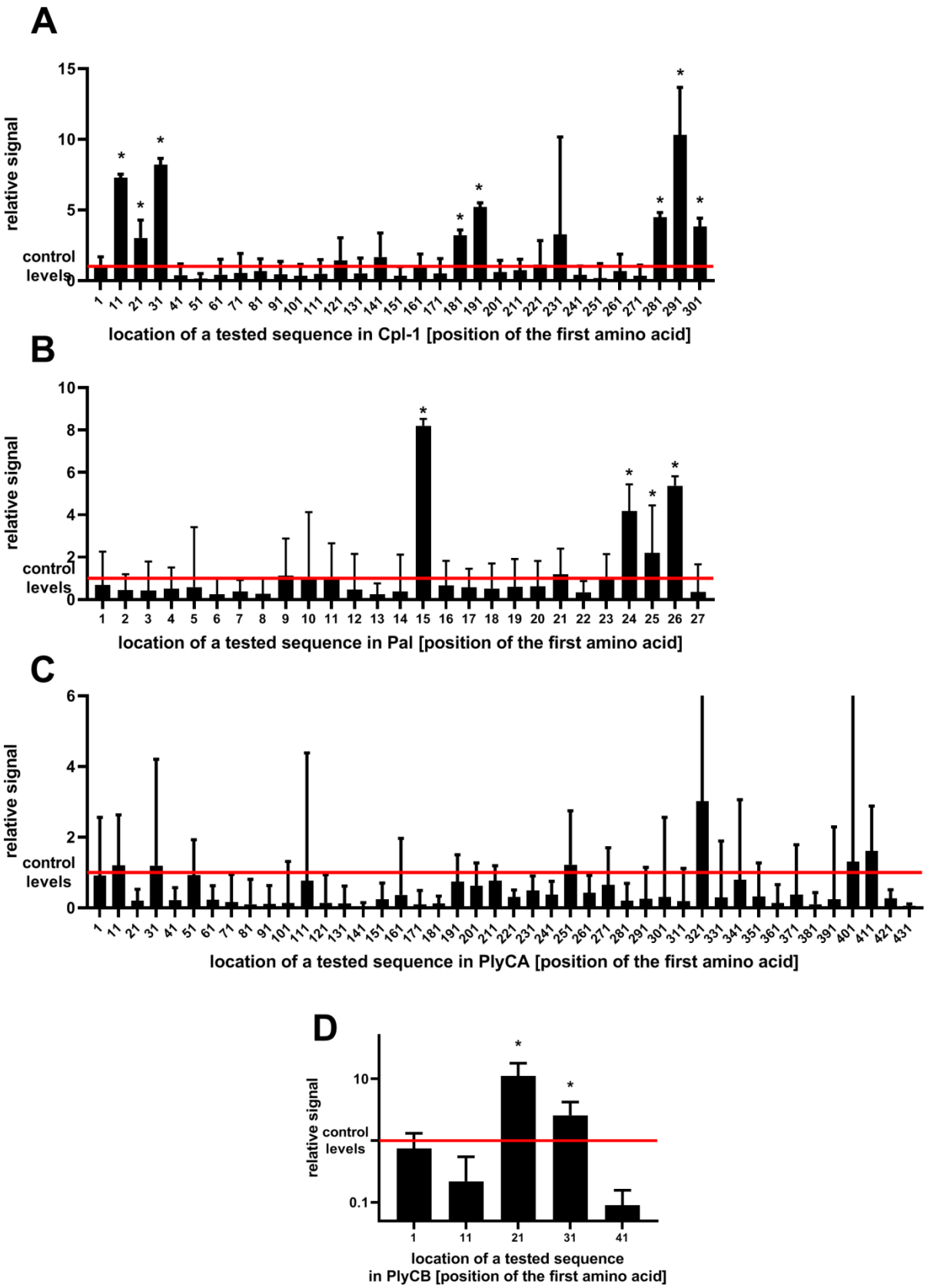

A
relative signal
control levels
location of a tested sequence in Cpl-1 [position of the first amino acid]
B
relative signal
control levels
location of a tested sequence in Pal [position of the first amino acid]
C
relative signal
control levels
location of a tested sequence in PlyCA [position of the first amino acid]
D
relative signal
control levels
location of a tested sequence in PlyCB [position of the first amino acid]

Figure 9: Immunogenic regions in endolysins Cpl-1 (A) Pal (B), PlyCA (C) and PlyCB (D) identified with human sera. Reactivity of human sera with phage-display library containing oligopeptides coding for fragments tailing endolysin sequence (10 aa shift) has been presented. Sera from healthy humans was used in immunoprecipitation. NGS sequencing was used to identify the immunoprecipitated fraction of the phage display library. The relative signal of oligopeptides after immunoprecipitation is presented, indicating the relative strength of interaction between 56 aa oligopeptide and specific IgG. Red lines represent the signal of the 'input sample' (that is: the library before immunoprecipitation), normalized as 1.0 * - adj. $p<0.05$ from comparison between normalized signals from immunoprecipitated sample and input sample, for each oligopeptide. Datapoints represent average and SD of 6 biological replicates (n=6), two technical replicates were completed.

3. Natural immunogenicity in humans is mediated by PlyCB subunit of PlyC protein

Human serum samples from healthy volunteers (n=56) were also tested for their reactivity to PlyC by ELISA. Anti-PlyC IgG levels showed significantly higher levels in 12.5 % of the samples (7 out of 56) compared to the overall population (outliers) (Figure 10A). Since the PlyC holoenzyme consists of two subunits, PlyCA (the domain of enzymatic activity) and PlyCB (the binding domain), both domains were tested individually. Approximately 10% (5 out of 56) of samples showed significantly higher anti-PlyCB IgG levels than the overall population. Noticeably, all 5 outliers for PlyCB layed within outliers detected for the whole PlyC holoenzyme. Levels of anti-PlyCA showed no outliers. Furthermore, anti-PlyCA IgG levels followed a normal Gaussian distribution whereas both PlyC and PlyCB significantly skewed from a Gaussian distribution because of overrepresentation of the high reads (Table 1). Next, reactivity of sera for PlyC, PlyCA, and PlyCB were plotted against each other to elucidate any correlations in reactivity within each individual. PlyC values were found to strongly correlate with those observed for PlyCB (Spearman $r = 0.820$, $p<0.0001$), however, correlation of PlyC with PlyCA was markedly lower and found to be statistically insignif-

icant (Spearman r = 0.24, p>0.01) (Figure 10B and 10C, Table 2). Taken together, the data suggests the PlyCB subunit mediates PlyC-IgG interactions in the investigated human sera.

Table 1. Evaluation of anti-PlyC, anti-PlyCA, and anti-PlyCB IgG levels in human sera by normal distribution analysis. The Anderson-Darling analysis tests the null-hypothesis that a cumulative distribution curve does not vary from expected Gaussian distribution. Similarly, the D'Agostina & Pearson analysis tests whether skewness and kurtosis of the data vary significantly from expected in Gaussian distribution.

	PlyCB	**PlyCA**	**PlyC**
Anderson-Darling test			
A2	3.391	0.5596	4.376
p-value	<0.0001	0.1415	<0.0001
Passed normality test (alpha=0.05)	No	Yes	No
D'Agostino & Pearson test			
K2	33.64	5.251	34.13
p-value	<0.0001	0.0724	<0.0001
Passed normality test (alpha=0.05)	No	Yes	No

Table 2. Correlations of human serum IgG reactivity to PlyC and its domains: PlyC vs. PlyCA (first row), and PlyC vs. PlyCB (second row).

	Spearman r	**p-values**	**CI**	**Pearson r**	**p-values**	**CI**
PlyCA	0.24	0.078	-0.035 to 0.48	0.32	0.017	0.06 to 0.54
PlyCB	0.82	<0.0001	0.71 to 0.89	0.95	4.9E-30	0.92 to 0.97

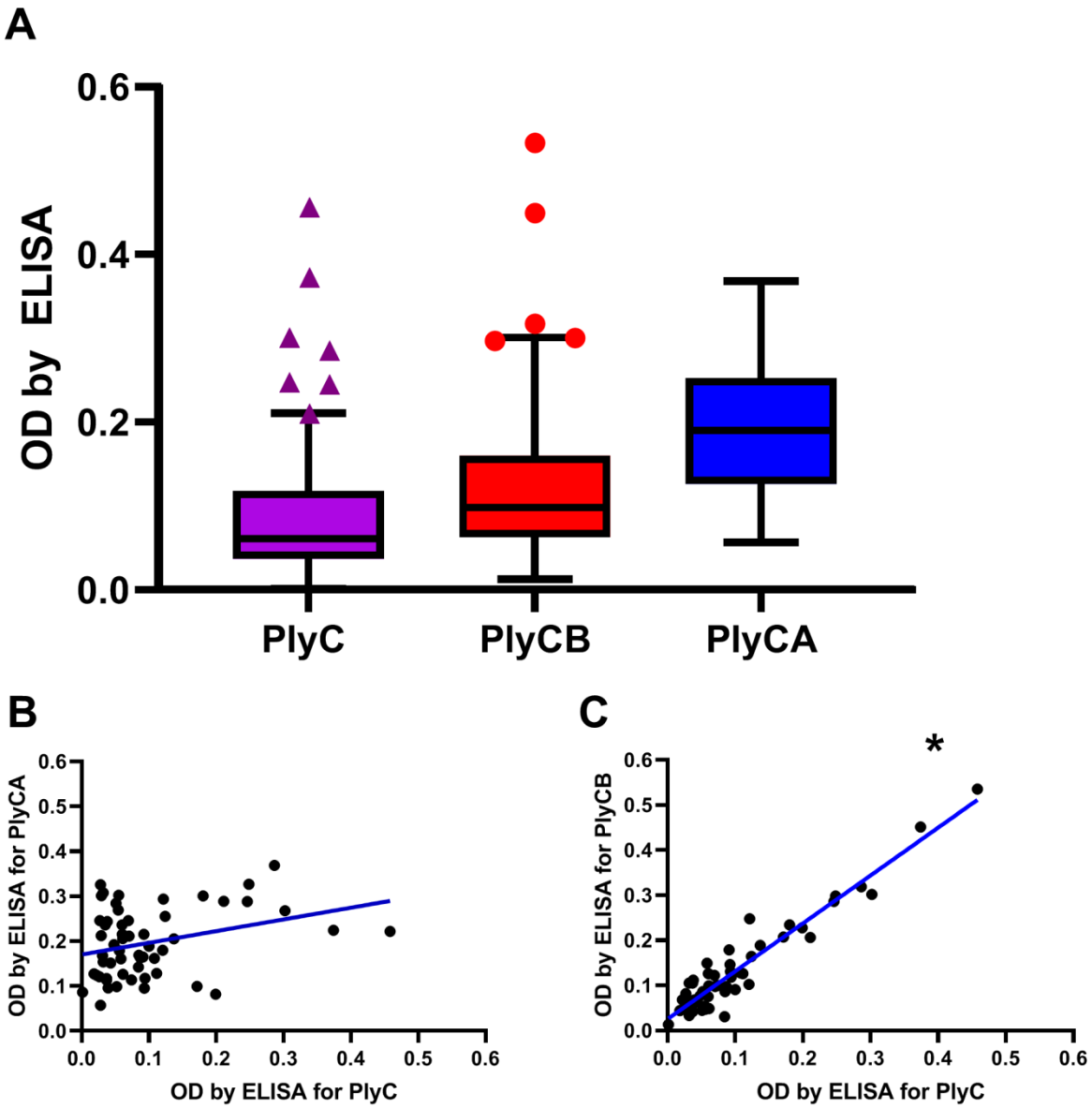


Figure 10. IgG reactivity against PlyC and its subunits, PlyCA and PlyCB. (**A**) Sera from healthy volunteers (n=56) were tested against recombinant PlyC, PlyCA, and PlyCB by ELIS. Levels of antibodies as ELISA signal; the mean value was presented as the vertical line, 25th and 75th percentile were presented as the box, whiskers designate outliers cut-off (Tukey method, GraphPadPrism), points (triangles or circles) represent outliers in each group. All five outliers that demonstrated the highest levels of anti-PlyCB IgG (circles) demonstrated also the highest levels of anti-PlyC IgG's (triangles). (**B**) Correlation (blue line) between levels of antibodies specific to PlyC and antibodies specific to PlyCA (insignificant). (**C**) Correlation (blue line) between levels of anti-

bodies specific to PlyC and antibodies specific to PlyCB. Correlation between levels of antibodies specific for PlyC and PlyCB was statistically significant. * - p-value < 0.0001.

4. Determination of immunogenic amino acids in Cpl-1 and Pal endolysins

Basing on the statistical analysis, selected oligopeptides were further processed for identification of amino acids that are directly engaged in endolysin-antibody interaction. These oligopeptides were located at C-end of Cpl-1 endolysin (301-339 aa) and within Pal endolysin (141-194 and 251-304 aa) (Figure 11). The oligopeptides were further analysed by site-directed mutagenesis. Consecutive pair or trio of amino acid in their sequences was substituted with alanine (or glycine, if a wild type amino acid was alanine). This yielded a mutagenized library that was tested with the same technology (Figure 7) and allowed to identify amino acids which substitution caused the loss of reactivity between an oligopeptide and endolysin-specific antibodies. Noteworthy, no oligopeptides showed statistically significant increase in enrichment what confirms immunogenicity of wild type oligopeptide.

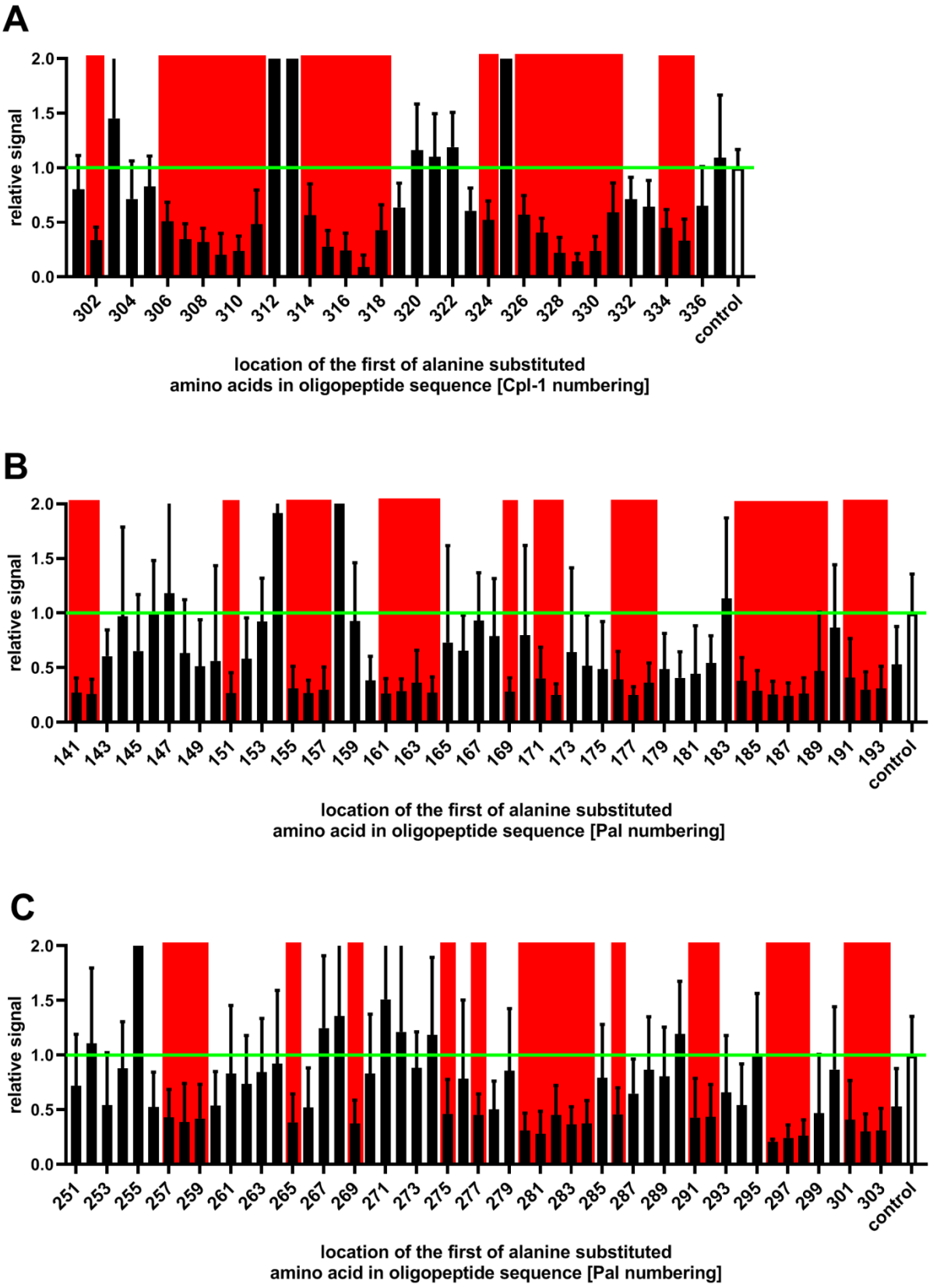

A
relative signal
2.0
1.5
1.0
0.5
0.0
302 304 306 308 310 312 314 316 318 320 322 324 326 328 330 332 334 336 control
location of the first of alanine substituted
amino acids in oligopeptide sequence [Cpl-1 numbering]
B
relative signal
2.0
1.5
1.0
0.5
0.0
141 143 145 147 149 151 153 155 157 159 161 163 165 167 169 171 173 175 177 179 181 183 185 187 189 191 193 control
location of the first of alanine substituted
amino acid in oligopeptide sequence [Pal numbering]
C
relative signal
2.0
1.5
1.0
0.5
0.0
251 253 255 257 259 261 263 265 267 269 271 273 275 277 279 281 283 285 287 289 291 293 295 297 299 301 303 control
location of the first of alanine substituted
amino acid in oligopeptide sequence [Pal numbering]

Figure 11: Amino acids that mediate immunogenicity of endolysins Cpl-1, Pal, and PlyC. Immunogenicity of C-end fragment of Cpl-1 (**A**), 141-193 aa of Pal (**B**) and 251-303 aa fragment of Pal (**C**) were defined in the reaction of a phage-display library consisting of oligopeptides with alanine substituted consecutive pairs or triplets (alanines were substituted with glycine). After incubation with murine sera exposed to a tested protein alanine substituted oligopeptides not recognized by specific to wild-type oligopeptides antibodies were washed away. Relative signal shows strength of recognition of this substituted oligopeptide. A low relative signal correlates with lack of recognition by wild-type antibodies meaning that immunogenic amino acids crucial for interaction with IgG antibodies have been substituted. Green lines represent the signal of the original, non-substituted oligopeptide, normalized as 1.0 red areas - adj. $p<0.05$ from comparison between normalized signals from immunoprecipitated sample and input sample for aminoacids whose substitution showed lowering of the relative signal for all oligopeptides with substituted aminoacid on this position. Datapoints represent average and standard deviation of 6 biological replicates (n=6), two technical replicates were completed.

4. Preclinical basic safety evaluation

1. Pal and Cpl-1 endolysins show no intrinsically toxic activity

In addition to specific immune reaction (IgG induction), biological drugs may potentially mediate other undesirable effects that should be excluded at advanced level of synthetic drugs design. Since my goal was to propose new variants of endolysins with improved performance *in vivo*, I assessed these enzymes for other (than IgG induction) undesirable effects. For this purpose, I conducted basic toxicity tests of selected endolysins (Cpl-1, Pal) to evaluate potential changes in mammalian gene expression profiles, changes in activity of a complement system, hypersensitivity (identified by IgE induction and overall health condition in mice treated with large quantities of the tested proteins).

Possible detection of increase in protein-specific IgE is a marker of hypersensitivity, since IgE fractions mediate allergic reactions. Here, no tendency for the development of Cpl-1- or Pal-specific

IgE was observed in mice (Figure 12). Serum levels of endolysin-specific IgE were significantly lower than in a positive control group and not significantly higher than negative control.

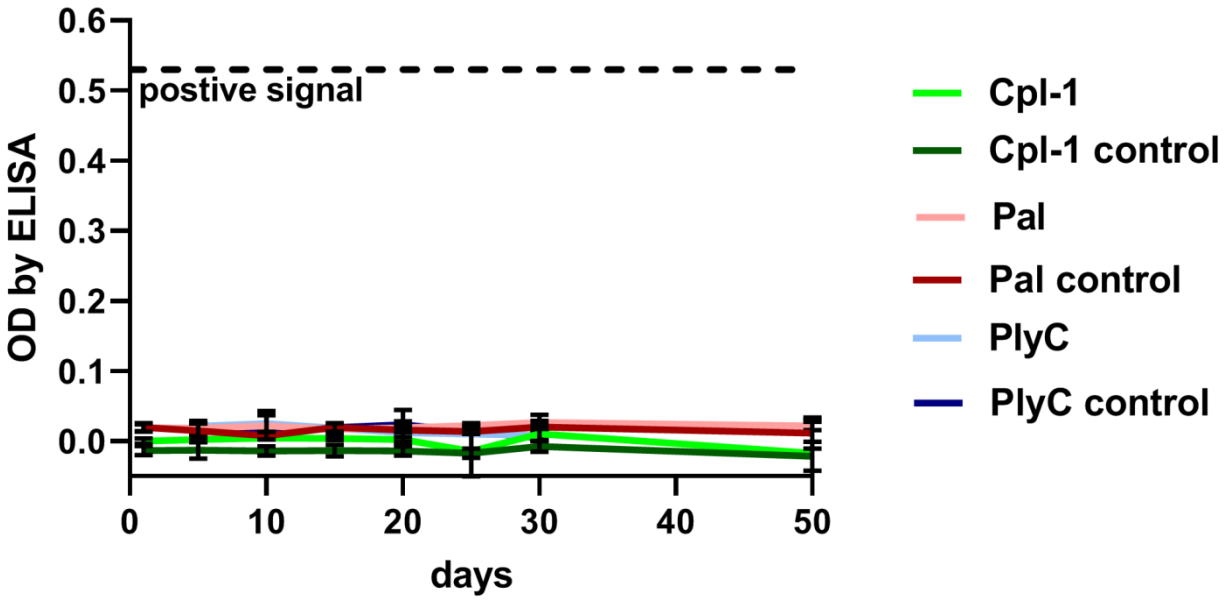


Figure 12: Antibody levels of IgE in mice challenged with Cpl-1, Pal and PlyC. Animals were exposed to purified protein by injection (i.p.). Serum samples were collected on days 1, 5, 10, 15, 20, 25, 30, and 50 and tested for levels of specific to Cpl-1, Pal or PlyC antibodies by ELISA. Positive signal- reference level of IgE antibodies detected in mice allergic to a model antigen (OVA). No significant difference detected between Cpl-1 or Pal treated groups and a respective (negative) control. Difference between positive signal and all groups at all time points is statistically significant ($p>0.05$). Each line represent average level in a group of mice (n=6).

Analyzing profiles of gene expression in eukaryotic cells before and after an exposure to investigated agents is a powerful tool for identification of possible harmful effects. Analysis of changes that those agents cause in eukaryotic cells can identify potential toxicity, oncogenic activity, and other problems. Human cell lines, pharyngeal carcinoma (FaDu) and normal SC macrophage, were exposed to 0.5 µM of bacteriolytic agent (17.5 µg/mL of Pal and 20 µg/mL of Cpl-1) for 6h. Cellular

mRNA was isolated and analyzed by microarray technology. Statistical analysis of microarrays results showed no statistically significant ($p< 0.05$) changes in gene expression after exposure to Cpl-1 or Pal (in comparison to those exposed to BSA as a control) in both cell lines. I performed two types of statistical analysis, the Bonferroni Holm method for the family-wise error rate (FWER) and the Benjamini Hochberg method for the false discovery rate (FDR). Results demonstrate that Pal and Cpl-1 has no negative effects on the human cells, including a toxicity, inflammatory response induction, or activation of oncogenes.

2. Pal and Cpl-1 do not activate complement system

The complement cascade/system is a component of blood playing a crucial role in development of immune responses. It enhances the ability of antibodies and phagocytic cells to remove damaged or abnormal cells or many types of microbes. Proteins that affect and interfere with this cascade can cause a serious harm during treatment, especially treatment of microbial infections. Commercially available test quantifies activity of classic, alternative, and MBL-dependent pathways. Pal and Cpl-1 endolysins were tested *ex vivo* (human blood) with Cpl-1 and Pal and compared with a control consisting of PBS only.

Cpl-1 and Pal in concentration of 2 μg/mL during testing showed no effect on the activation of the complement system (classic, alternative or MBL-dependent pathway) compared to the control with PBS only ($p <0.05$) (Figure 13). In other words, Pal and Cpl-1 do not affect significantly first line of non-cellular immune response in humans.

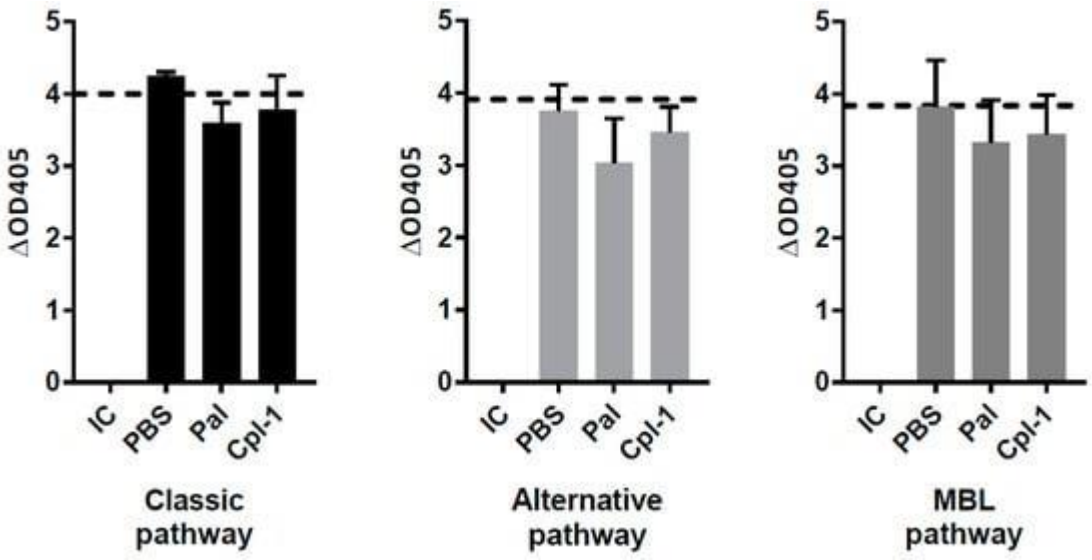


Figure 13: Levels of classic, alternative and MBL pathway of the complement system in human serum after exposure to a concentration of 2 µg/mL. IC: inactivated complement system (as supplied by the manufacturer), dashed line - control of normal complement activity (as supplied by the manufacturer).

3. Mice challenged with Cpl-1 and Pal show no adverse side effects and no inflammatory response after treatment with large amounts of the proteins.

Evaluation of general health condition in animal models and side effects of potential therapeutics may demonstrate adverse activity that cannot be identified *in vitro* and *ex vivo*. For this reason, a general assessment of mice treated with Cpl-1 or Pal endolysins was performed. The health status scoring was used after 2 h and 5 h after exposure to Cpl-1 or Pal. Since in the endolysin preparation some trace amounts of LPS was detected, negative controls were treated with the same level of LPS as contained in the purified endolysin solutions. A toxic dose of LPS (2000 EU per mouse) served as a positive control of toxicity. An inflammatory marker (IL-6) was also assessed in these animals.

Cpl-1 and Pal exerted no negative effects 2h and 5h after exposure to 0.3 mg of endolysin (Figure 14). No significant difference in health status scorings was detected in the negative control mice and endolysin treated mice ($p<0.05$). Positive control (mice exposed to high LPS dose) showed a marked, negative effect at both time points and significantly worse health status than animals treated with the endolysins ($p<0.05$). Measurements of IL-6 levels in murine blood 5 hours after treatment with the endolysins showed no significant differences to negative control mice ($p>0.05$) (Figure 15). Both animals treated with endolysins and negative controls showed significantly lower levels of IL-6 than the positive control consisting of mice exposed to high LPS dose ($p<0.05$).

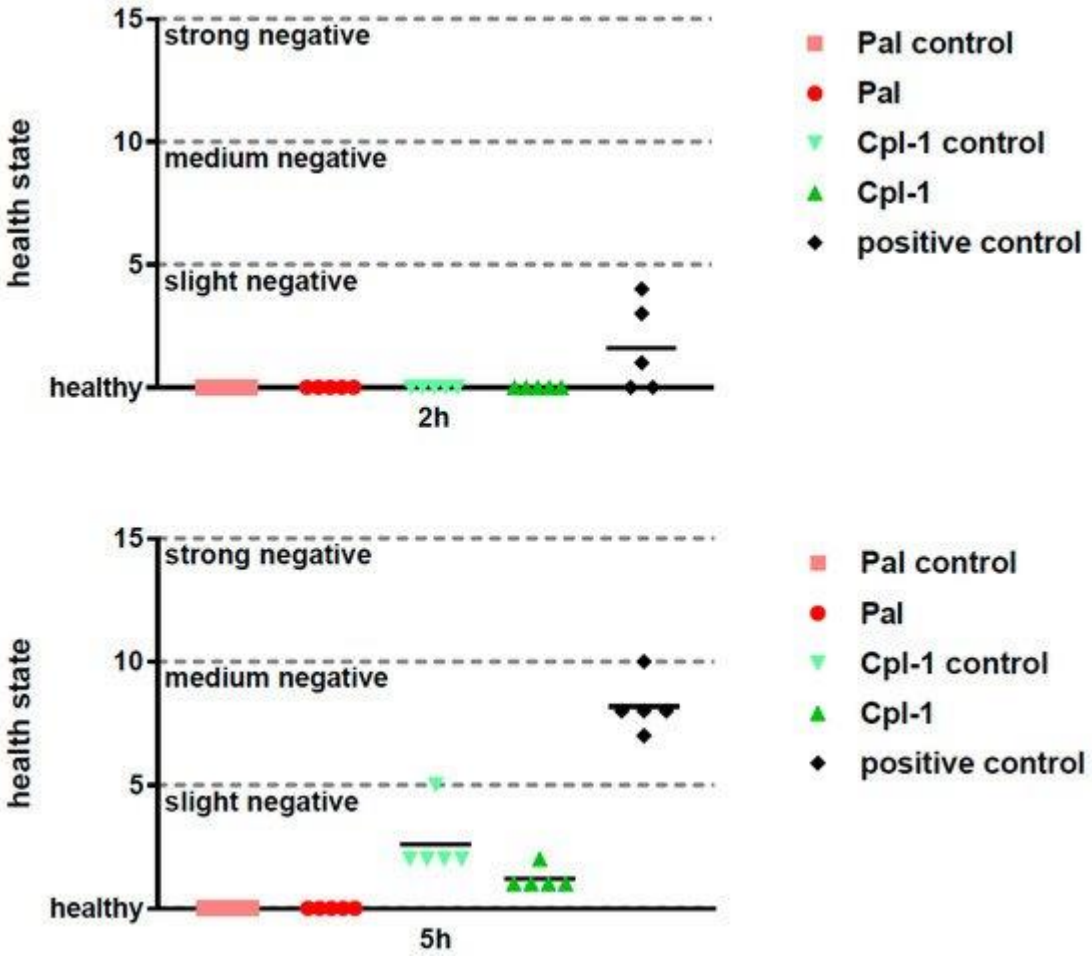


Figure 14: Health status scoring in mice treated intraperitoneally with Pal and Cpl-1. Pal: mice treated with 0.3 mg of Pal; Pal control: mice treated with PBS containing the same residual LPS content as Pal (0.6 EU); Cpl-1: mice treated with 0.3 mg of Cpl-1; Cpl-1 control: mice treated with PBS containing the same residual LPS content as Cpl-1 (8 EU); positive control: mice treated with a toxic LPS dose (2000 EU). Upper panel: 2 h after treatment; Lower panel: 5 h after treatment.

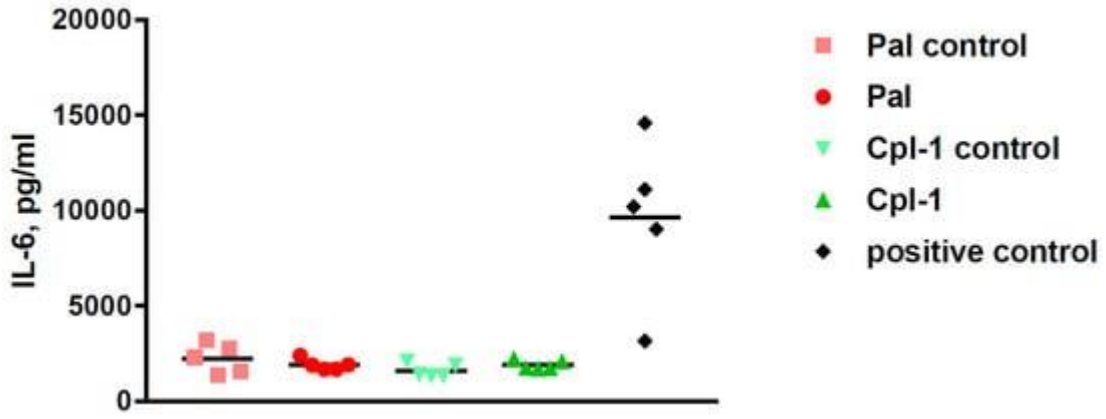


Figure 15: IL-6 cytokine levels in murine blood 5 h after intraperitoneal treatment with Pal or Cpl-1. Pal: mice treated with 0.3 mg of Pal; Pal control: mice treated with PBS containing the same residual LPS content as Pal (0.6 EU); Cpl-1: mice treated with 0.3 mg of Cpl-1; Cpl-1 control: mice treated with PBS containing the same residual LPS content as Cpl-1 (8 EU); positive control: mice treated with a toxic LPS dose (2000 EU).

5. Deimmunization of endolysins Pal and Cpl-1

Since other types of analyses did not reveal any important biological activities of Cpl-1 and Pal, my further work was focused on decreasing their ability to enter interactions with IgG antibodies. Basing on my previous results: identified amino acids that mediate interaction of immunogenic regions in the investigated endolysins with IgGs, and having developed a sensitive and precise method for testing bacterial lysis efficacy (fluorometric assay), I conducted deimmunization of selected endolysins. Deimmunization was conducted in Pal and Cpl-1, since identification of their immunogenic regions gave consistent results in animal model and in human population. Thus, potential for deimmunization had both good technological basics and high applicability. PlyC endolysin was excluded from further studies due to differences in reactions between human sera that recognized PlyCB subunit as immunogenic and murine that recognized PlyCA subunit as an immunogenic one.

Engineering of Pal and Cpl-1 to escape recognition by epitope-specific antibodies was conducted by amino acid substitutions with the following assumptions:

- different charge and chemical type of substituted aminoacids prevents establishing the same interaction between IgG and antigen;
- lowering of folding energy (ddG) with substitution improves stability and creates minimal risk of severe conformational changes
- larger aminoacids are avoided as possible in order to avoid creation of additional steric tensions.

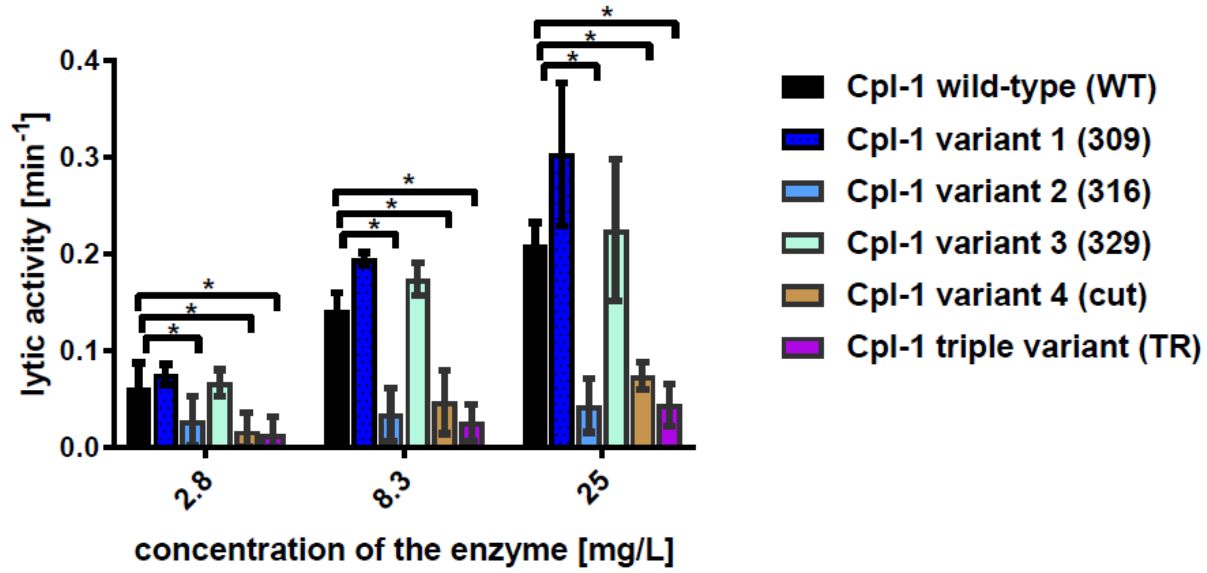
Nine possible variants of Pal were designed (Table 3) and ORF coding these variants were synthesized, cloned and expressed. Expression in an *E. coli* expression system yielded three variants that expressed as soluble proteins: Pal v1, Pal v3, Pal v9. Five Cpl-1 (v1 -v5) variants were designed and processed similarly as Pal variants. They all yielded satisfactory expression of soluble enzymes (Table 3).

1. Pal and Cpl-1 variants with engineered immunoreactive epitopes show altered bacteriolytic efficiency

All efficiently expressed variants of Pal and Cpl-1 were tested for their antibacterial activity proving that all engineered and expressed endolysins are still capable to lyse target *Streptococcus pneumoniae*. Activity was tasted in a range of enzyme concentrations and for v1 and v9 of Pal and variants v2, v4, and v5 of Cpl-1 was significantly lower than in the wild-type enzymes (Figure 16). Variants Cpl-1 v1 and Cpl-1 v3 were active similarly to WT enzyme, and Pal v3 was significantly more active against sensitive bacteria than the wild-type enzyme (Figure 16). Since all variants demonstrated at least partial antibacterial activity, they were all used in further immunological studies. Every variant of Cpl-1 endolysin with modified 316 aa lost majority of its bacteriolytic activity

suggesting this amino acid to be important for stability of a catalytic domain in spite of being far from an active site in primary structure. Difference between variants Pal v1 and Pal v9 is one amino acid but their activity is significantly different.

A


lytic activity [min-1]
0.4
0.3
0.2
0.1
0.0
2.8
8.3
25
concentration of the enzyme [mg/L]
Cpl-1 wild-type (WT)
Cpl-1 variant 1 (309)
Cpl-1 variant 2 (316)
Cpl-1 variant 3 (329)
Cpl-1 variant 4 (cut)
Cpl-1 triple variant (TR)

B

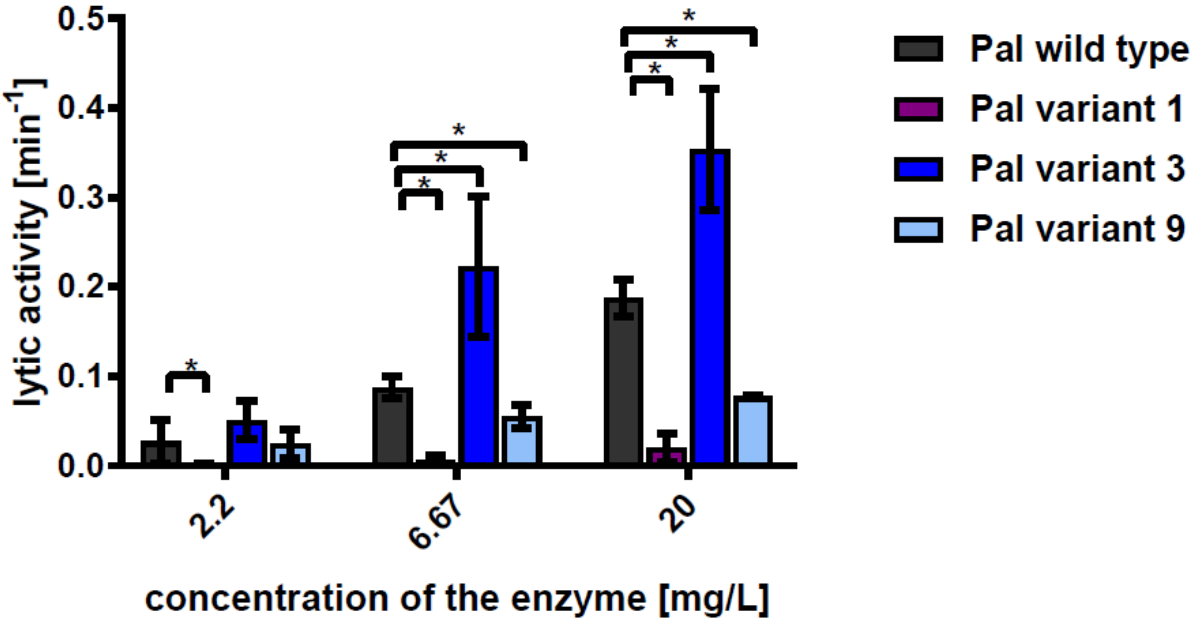

lytic activity [min-1]
0.5
0.4
0.3
0.2
0.1
0.0
2.2
6.67
20
concentration of the enzyme [mg/L]
Pal wild type
Pal variant 1
Pal variant 3
Pal variant 9

Figure 16 Antibacterial activity of endolysins Activity of Cpl-1 (A), Pal (B), and variants with engineered epitopes identified by EndoScan. Activity was measured in PBS. * - adj. $p<0.05$. Bars represent an average, two technical replicates were completed. Whiskers represent standard error of the activity.

Table 3. Designed Pal and Cpl-1 variants with engineered immunoreactive epitopes.

protein	variant designation	position of a modification	wild-type sequence	modified sequence
Pal	v1	257-259	MKS	TFK
	v2	265-268	YNDG	DKGL
	v3	280-282	DKP	GGA
	v4	283-285	QFT	GLA
	v5	257-259 265-268 280-285	MKS YNDG DKPQFT	AAA AAAA AAAAAA
	v6	257-259 265-268 280-285	MKS YNDG DKPQFT	TFK DKGL GGAGLA
	V7	257-259 280-282	MKS DKP	TFK GGA
	v8	265-268	YNDG	LAGL
	v9	257-259	MKS	TFG
Cpl-1	v1	309-311	SGK	DAW

	v2	316	M	R
	v3	329-331	TKE	LQP
	v4	309-311 316-317 329-331	MKS MN TKE	DAW RF LQP
	v5	309-339	SGK…TVA	deleted

Biologically active proteins can be prone to non-specific inactivation by the complex serum matrix. This might change the variants' applicability *in vivo*, independent of their possible improved performance in the presence of specific antibodies. Endolysin variants were therefore tested for their antibacterial activity *ex vivo* in naïve sera (from unchallenged animals). In wild-type enzymes and in most of the tested variants similar, significantly decreased activity was observed (Figure 17). Two Cpl-1 variants, Cpl-1 v4 and Cpl-1 v2, demonstrated no significant change in their activity with or without serum, but overall activity of these variants was poor in general. Pal v1, also with intrinsic poor activity, demonstrated improved activity in the presence of serum, but it was still lower than that of the WT protein (Figure 17).

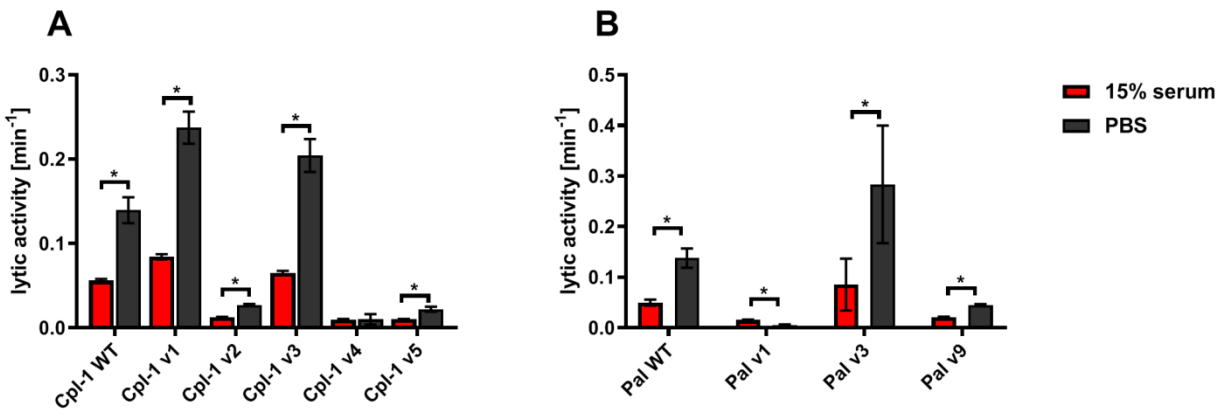


Figure 17. Effect of naïve (non-specific) murine serum on antibacterial activities of endolysins Pal (A) and Cpl-1 (B), and variants with engineered epitopes identified by EndoScan. Concentration of endolysins and all variants was 10 μg/mL. Serum concentration was 15%. Non-inactivated serum was used. * - $p<0.05$. Bars represent average, two technical replicates were completed. Whiskers represent standard error of the activity.

2. Ability of new Pal and Cpl-1 variants to induce specific IgG

Pal and Cpl-1 variants were further investigated for their *in vivo* immunogenicity in an animal model. Immunogenicity herein was considered as efficiency of a protein to induce specific antibodies (measured as serum levels of specific IgG). None of the variants demonstrated significantly changed immunogenicity to Cpl-1, but in the case of Pal v9 its immunogenicity was significantly decreased (only on days 40 and 50 after challenge, $p<0.001$) and for Pal v1 its immunogenicity was significantly increased (only on days 40 and 50 after challenge, $p<0.001$) (Figure 18). Thus, the overall ability of the variants to induce specific antibodies in comparison to wild-type enzymes was moderately decreased only in one case, and this variant demonstrated lower antibacterial activity than the WT endolysin.

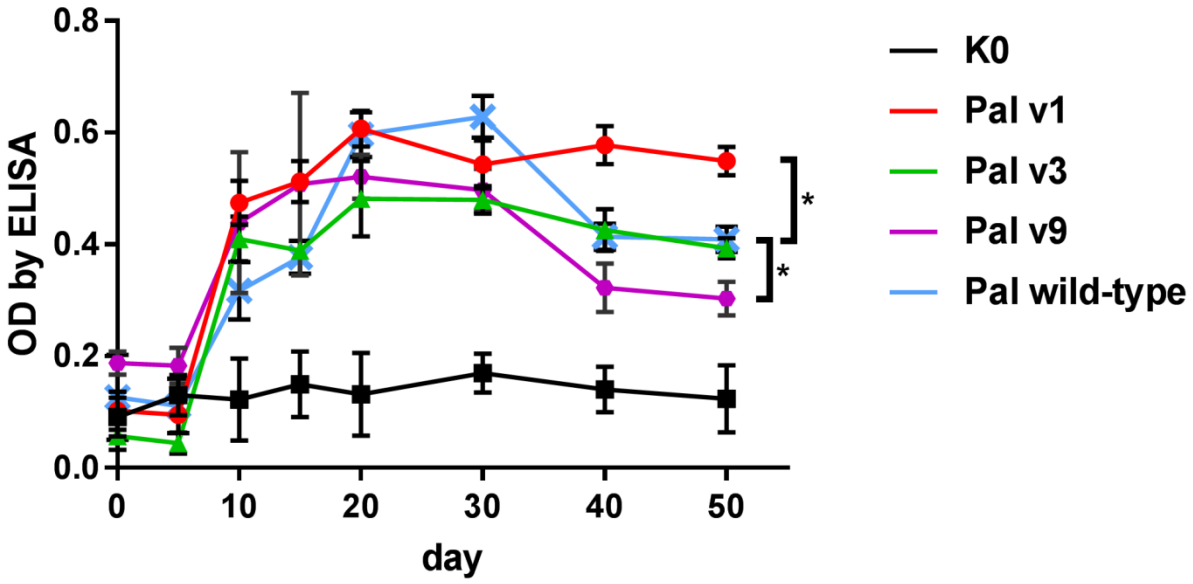


Figure 18. Serum levels of specific IgG induced by Pal and its variants with engineered epitopes identified by EndoScan. Levels of specific IgG antibodies detected by ELISA. Mice (n=6) were exposed (i.p.) to variants of Pal endolysin (100 µg/mouse), Pal endolysin (wild type) and PBS (K0). Sera were collected until 50th day after initial exposure and levels of specific to a respective protein were measured. * - p value < 0.05 between a group exposed to variant and wild type Pal. Each group represent six mice (n=6), two technical replicates of each sample were conducted, points represent average and whiskers represent standard deviation.

3. Ability of new variants of Cpl-1 and Pal to escape cross-reactions with IgGs induced by the wild type endolysins

Specific antibodies induced by an active protein can potentially limit success of this protein in situations of repeated use. Protein variants with the same overall immunogenicity but different immunogenic epitopes may escape cross-reactions with antibodies induced by wild-type protein, thus being useful in repeat treatments. Therefore, we investigated cross-reactions of Pal and Cpl-1 variants with IgG induced by wild-type endolysins. In all variants, cross-reactivity was weaker than the reactivity of WT Pal or Cpl-1, with the decrease of the Pal variants' cross-reactivity being more marked

(Figure 19). Therefore, we next tested *ex vivo* how the Pal variants escaped cross-neutralization by specific serum induced by WT Pal. Lytic activity of Pal WT, and its variants v3 and v9, was compared in the reaction with Pal-specific serum used as a blocking agent. Pal v1 was not investigated due to its overall significantly weakened antibacterial activity. This testing revealed that Pal v3 demonstrated significantly stronger antibacterial activity in Pal-specific serum than the WT Pal enzyme ($p<0.05$). Specifically, two effects contributed to this advantage: Pal v3 was not neutralized by Pal-specific serum, and it demonstrated intrinsic higher activity than WT Pal (Figure 20). Thus, Pal v3 is a variant of the Pal endolysin that escapes cross-neutralization with specific antibodies induced by the WT Pal, and it has overall improved antibacterial activity.

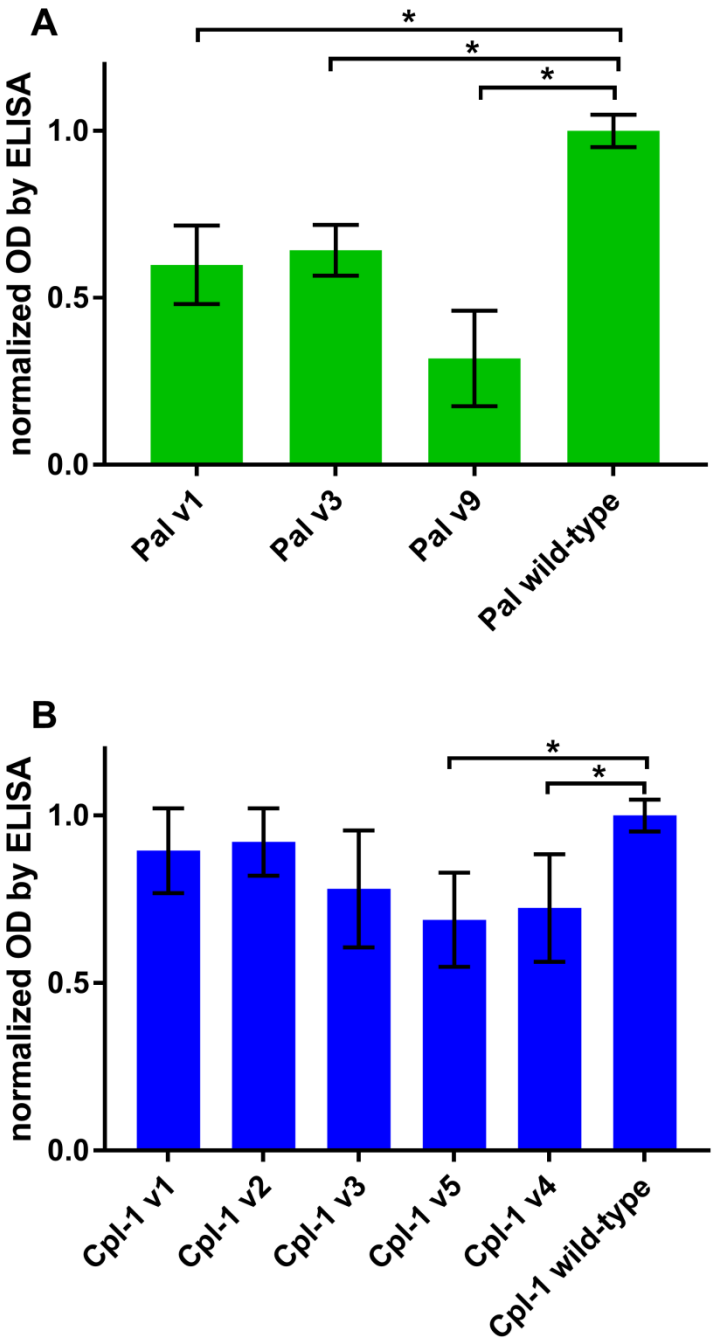


Figure 19. **Cross reactivity of Cpl-1, Pal and their variants by specific to wild type murine serum.** Comparison of immunogenicity (cross-reactivity) of Pal (**A**) and Cpl-1 (**B**) variants with engineered epitopes identified by EndoScan to wild-type enzyme specific IgG. * - adj. p<0.05. Bars show average, normalized levels of ELISA from six murine sera (n=6), whiskers show standard deviation. Two technical replicates of this experiment were performed.

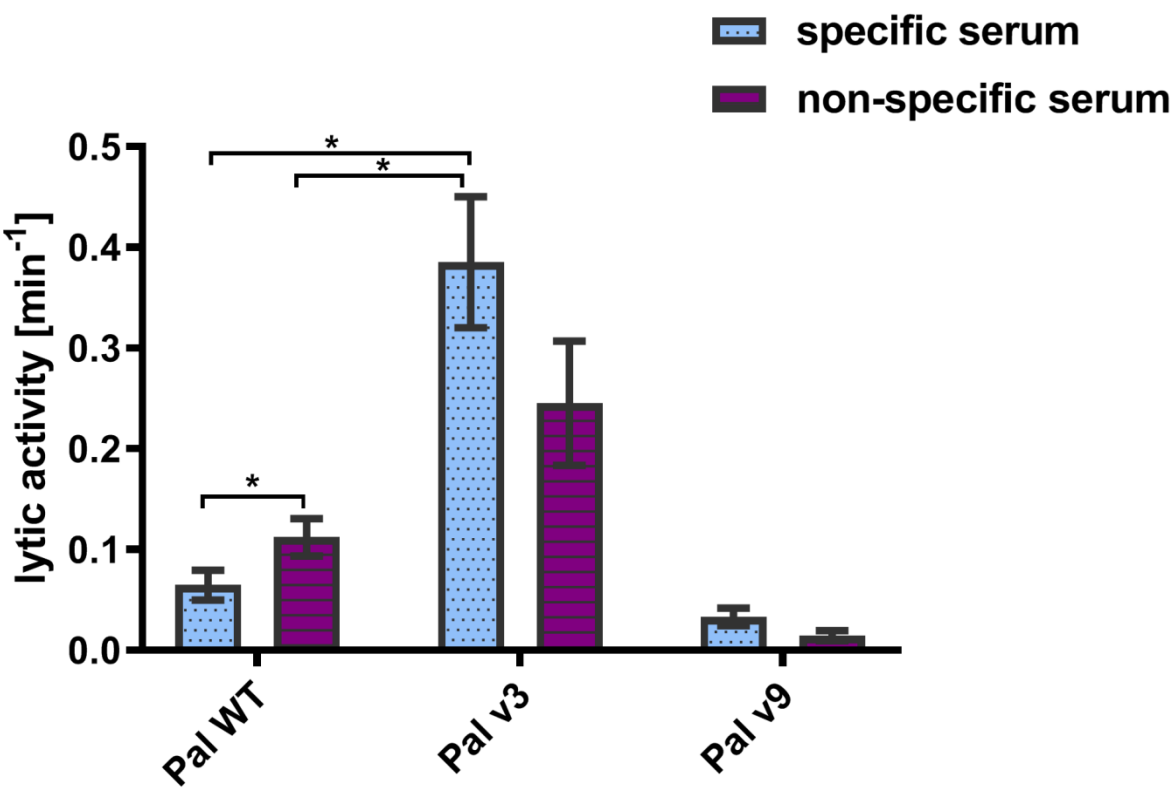


Figure 20. Bacteriolytic activity of Pal and Palv3 endolysins in serum specific to Pal WT. Bacteriolytic activity was measured in murine serum with confirmed high titer of IgG specific to Pal WT and compared with serum without detectable level of such IgG. Concentration of endolysins is 20 μg/mL. Serum concentration was 15%. * - $p<0.05$. Bars show average, two technical replicates were completed. Whiskers represent standard error of the activity.

10. Discussion

Bacteriophage endolysins gain increasing attention. These enzymes are promising and potent alternatives to traditional antibiotics. In the past decade at least three companies have been enrolling patients for Phase 2 clinical trials with endolysin-based therapeutics, according to **clinicaltrials.gov**. SAL200, a pharmaceutical composition containing the SAL-1 endolysin, which targets *Staphylococcus aureus* showed no toxicity in rodent and dog single- and repeated-dose studies and pharmacology studies demonstrated no signs of toxicity in central nervous and respiratory system function tests(70,71). Next, SAL200 was tested in monkeys to reveal pharmacokinetic and safety information. No laboratory abnormalities or adverse events were detected after injection of a single dose (up to 80 mg/kg per day) or injection of multiple doses (up to 40 mg/kg per day) and finally, the results for two human Phase 1 trials of SAL200, the first in-human studies for a bacteriophage endolysin-based drug, were recently published(71). SAL200 was well-tolerated, and no serious adverse effects were observed in this clinical study, with most adverse events being mild, self-limiting, and transient. Furthermore, no clinically significant values with respect to clinical chemistry, hematology, coagulation, urinalysis, vital signs, or physical examinations were observed.

In a study presented in this work, gene expression analyses of human SC macrophages and human pharyngeal FaDu cells after exposure to Pal and Cpl-1 are presented. Macrophages were chosen as they allow for the identification of potential effects of the endolysins on immune responses, whereas the pharyngeal cells represent non-immunological cells from a body site commonly colonized by *S. pneumoniae*. In both types of cells, no statistically significant changes in the expression levels of over 16,000 genes were noted compared to albumin-treated cells indicating that no specific pathways, like apoptosis, oncogene or inflammatory responses, were activated by the enzymes, thus strongly supporting the safety of Pal and Cpl-1. It should be noted that Entenza et al. (72) found

increased serum pro-inflammatory cytokine (i.e., IL-1β, IL-6, TNF-α, and INF-γ) levels in a rat endocarditis model of *S. pneumoniae* when treated with high, continuous levels of Cpl-1. However, both our gene profiling results as well as the direct serum measurements of IL-6 showed no increase due to endolysins, suggesting that the results of Entenza et al. reflect responses to the lysed bacterial cells rather than endolysins themselves. This is supported by Witzenrath et al. (73), who showed that lower levels of Cpl-1 reduced pro-inflammatory cytokines in infected mice in comparison to untreated infected mice. It was hypothesized that higher concentrations of Cpl-1 generate a more complete digestion of the peptidoglycan, thereby generating more pro-inflammatory fragments to induce the response. Due to the lytic nature of these enzymes, detailed dosing studies in conjunction with safety profiles will be needed to achieve an optimal safe, effective dose. Nonetheless, our results demonstrate that the purified enzymes themselves cause no inflammatory responses.

General prove of preclinical tests showed safety of both endolysins in vivo, at concentrations (0.3 mg/mouse; ~15 mg/kg based on a 20 g mouse) higher than previously used to show the efficacy of Cpl-1 in a pneumococcal sepsis model(74). Here, the overall health status scoring indicated no adverse effects of endolysin treatment. The endotoxin levels in the Cpl-1 preparations was >10 times higher than the levels in the Pal preparations (8 EU vs. 0.6 EU), demonstrating the importance of further optimization of endotoxin removal and testing. Moreover, our results demonstrate for the first time in any pneumococcal endolysin that they do not generate an IgE response, that would indicate association with hypersensitivity and allergic reactions. Still, active enzymes can be prone to non-specific inactivation by serum components.

In this work, I developed a new fluorimetric assay that employs a DNA dye, Sytox Green, for detection of bacterial lysis I validated this new method with two phage endolysins, Cpl-1 and Pal, both bacteriolytic toward *Streptococcus pneumoniae*, in varying concentrations of enzymes and bacteria.

This new method for the detection of bacterial lysis and for characterization of bacteriolytic agents provides a unachievable earlier, very high sensitivity and it is able to yield a kinetics curve even at concentrations of bacteria as low as 3.4×10^3 CFU/mL (6.8×10^2 CFU/sample). Importantly, the method offers direct, real-time measurements of reaction kinetics, thus allowing for calculation of progress and lytic activity parameters. A mathematical approach to evaluate lytic activity measured from raw reads has been proposed and it has demonstrated fluorescence dyes applicability in varying concentrations of two endolysins and bacteria. To demonstrate the potential of Sytox Green, I compared it to commonly used turbidity reduction assays thatdetect absorbance of 600 nm light passing through a sample whereby a drop in the absorbance is considered an indirect indicator of bacterial lysis. As demonstrated in this study, the use of DNA dye allows for detection of higher lytic activity of both enzymes, is more responsive, efficient, and capable for measuring progress of bacterial lysis in very low bacterial concentrations with no requirement of bacterial growth, thus having potential for detection of differences non-detectable for indirect turbidity assays like TRA and testing of samples collected without further processing. I also propose the use of maximum specific activity for general comparisons of different endolysins.

Major limitations for this fluorometric assay can be estimated from presented data. In the tested conditions, activity can be detected in enzyme concentration as low as 0.36 mg/L for Cpl-1 and 0.1 mg/L for Pal. The maximum detectable activity was over 1.0 min^{-1} for the Pal enzyme, which can be interpreted as complete lysis in less than 2 min. The highest measured concentration of bacteria is over 8×10^7 CFU/mL (0.2 mL volume of the sample) and lowest concentration of bacteria is 3.4×10^3 CFU/mL (680 CFU in 0.2 mL volume of the sample). I hypothesize that lower concentration of bacteria can probably be measured by extending the monitoring time, but this problem requires further investigation. All these listed limits are, however, marked improvements in compari-

son to TRA. Additionally, a fit of a mathematical model to describe how the Pal enzyme activity (one-phase association) changes in response to different concentrations of bacteria has not presently been tested, but represents the next planned extension of this assay for my future experiments.

Implementation of this protocol in other laboratories requires determination of two parameters:

- concentration of bacteriolytic agent that lyse all bacteria in the sample during the experiment, typically with agar plating method (lysis control),

- concentration of fluorescent dye that is not saturated during lysis (positive control).

There are three technical requirements to implement the protocol described herein. If a signal from lysis control (1) shows initial increase, (2) reaches a plateau near the end of the assay and (3) signal from positive control is around 33% larger than the plateau, the fluorescent protocol can be implemented. Determination of optimal bacteria/bacteriolytic agent concentrations or time/interval of the measurements during the assay improves the reproducibility and precision of the measurements, but still is technically optional. Implementation of the mathematical model is highly beneficial for comparison between other lytic agents and different bacteria concentrations, but not mandatory.

Traditionally, application of Michaelis-Menten calculations for bacteriolytic enzymes has been challenging since the substrate for the endolysin is the peptidoglycan, yet, the measured signals represent either light scattering, or in our case, DNA bound by the fluorescent dye rather than lysed peptidoglycan (75,76). Michaelis-Menten calculations measure the correlation between concentrations of substrate and the rate of the enzymatic reaction. Moreover, a bacterial cell is not a typical substrate due to the capability to simultaneously bind thousands of endolysin particles despite only a single lytic event “converting” a cell into a signal. Nonetheless, calculations of a Michaelis-

Menten model for our data reveals a strong correlation between enzyme activity and a bacteria lysis, thus providing grounds for prediction of enzyme behavior in comparison to other enzymes, and supporting an earlier hypothesis that a fixed number of bonds in the peptidoglycan need to undergo reaction before the bacteria is lysed(77). In the opposite case, varying number of bonds required for the lysis would cause high randomness in the model and large errors (75). Of note, our use of mathematical models and calculations is limited to the data available in this study and it requires further testing and validation in other conditions, for instance, the one-phase association model used herein for Pal endolysin. Other attempts to characterize Michaelis-Menten for endolysin characteristics exist, such as the EnzChekTM Lysozyme Assay Kit (Invitrogen) that uses labelled purified peptidoglycan. This commercially available assay, however, does not target the phenomenon of bacteria lysis, instead it detects peptidoglycan-derived substrate hydrolysis by agents of strictly defined specificity (78).

Sytox Green produces its signal by interaction with released bacterial DNA, but it can also penetrate into metabolically inactive bacteria, so the fluorescent signal is directly correlated to the amount of bacteria that have been inactivated, not necessarily completely disintegrated (76). Minor damage of the peptidoglycan that renders bacteria inactive but not destroyed cannot be detected by the indirect turbidity reduction assay. Moreover, large particles (if present in the investigated solution) blocking light in the indirect turbidity assay may strongly affect results, but in the fluorometric assay, where the dye itself is the source of light, such particles may have a lesser impact. On the other hand, both fluorometric assay and turbidity reduction assay employ light detection. Thus, they both may have decreased precision in environments that are not fully transparent for specific spectra. Importantly, fluorescent DNA dyes are constantly being developed, including a very wide range of offered spectra (79–81).

This presented approach can be used for rapid detection of susceptible bacteria in a variety of sample matrices or even in a mixture of bacterial strains when a defined selective bacteriolytic agent (e.g., a specific endolysin) is used. Further, with the use of the well-defined specificity of a bacteriolytic agent, bacterial identification can be achieved. Since fluorescent dyes that bind DNA deliver quantitative, real-time measurable signal of bacterial lysis, they have a potential for development of new, precise, high-throughput, low-cost, and rapid tests, applicable even in very low concentration and/or mixtures of bacteria.

The World Health Organization (WHO) and the Centres for Disease Control and Prevention repeatedly called for development of new or alternative antibacterial treatments (World Health Organization (47,48,50,82,83). I propose the use of fluorescent dyes and this fluorometric assay as a way to develop and improve methods broadening possibilities in microbiology both in the research of bacteriolytic agents and in the analysis or identification of bacteria samples

My results show that specific IgG response is readily generated against endolysins, which is not unexpected due to the proteinaceous nature of these enzymes. I have identified immunogenic regions within the investigated endolysins and then within Cpl-1 and Pal I identified specific amino acids that are engaged in the endolysin-IgG interactions. This identification was used to design new variants of Cpl-1 and Pal with potentially decreased immunoreactivity. Intrinsic immunogenicity (an ability of a variant to induce specific IgG) was assessed in a mouse model (Figure 18) and significant differences were only noted for two Pal variants, v1 and v9, although in the case of v1 its intrinsic immunogenicity was higher than that of Pal WT. Only in animals challenged with Pal v9 was the induced level of enzyme-specific IgG lower than in animals challenged with Pal WT. Moreover, this effect was detectable only in long-term observation, that is on days 40 and 50 after challenge, but not at the earlier phase of response development (Figure 18). Thus, the effect of

deimmunization was very limited, since it was only observed in a later phase and the overall level of detected antibodies was significantly higher than that observed in non-challenged control mice. This suggests limited applicability of this particular variant as a deimmunized endolysin. Moreover, the Pal v9 variant demonstrated decreased intrinsic antibacterial activity (Figure 16) due to the amino acid substitution, so its potential for therapeutic use can be expected to be weak. Nonetheless, our observation demonstrates that antigenic epitopes identified by epitope scanning can be deimmunized by substitutions of amino acids identified as key elements of the epitopes.

I further investigated cross-reactions of Pal and Cpl-1 variants with IgGs induced by wild-type enzymes, since escaping cross-reactions may be useful in repeated treatments. In all Pal variants cross-reactivity was weaker than the reactivity of wild-type Pal. For Cpl-1, it was weaker in variants v4 and v5. Since the overall bacteriolytic activity of Cpl-1 v4, Cpl-1 v5 and Pal v1 was low, I continued with Pal v3 and v9 in an *ex vivo* assay by blocking with WT Pal-specific serum. This assay is a model of *in vivo* conditions in a human or animal organism where previous treatment/s with the WT endolysin yielded increased levels of specific IgG. No decrease of Pal v3 activity in WT Pal-specific serum in comparison to its activity in naïve specific serum was observed. Thus, this variant escapes cross-reaction and resulting cross-neutralization by specific antibodies induced by WT Pal. Furthermore, overall antibacterial activity in the presence of WT Pal-specific serum was significantly higher than of WT Pal in the same conditions. Two effects contributed to this advantage: Pal v3 was not neutralized by Pal-specific serum, and its intrinsic activity was higher than Pal WT.

Possible reason of low effectiveness of Cpl-1 deimmunization is the immunogenic region in the catalytic domain of the enzyme that was not modified (not to disrupt the enzymatic activity) and retained its immunogenic activity in all variants. Noteworthy, Cpl-1 v4 with all three proposed epitopes altered shows also altered immunogenicity. It can be hypothesized that epitopes in fact has

been detected, but strength of each signal epitope is too low in comparison to overall immunogenicity. Joined effect of not altered epitopes close to the active site and relatively weak effect of each single epitope tested cause alteration in immunogenicity to be visible only in variant with all three chosen epitopes altered (Cpl-1 v4) or with all three epitopes removed (Cpl-1 v5).

I observed that in all Pal variants reaction to WT Pal-specific serum was weaker than that of WT Pal, and in Cpl-1 it was weaker in variants v4 and v5. Overall bacteriolytic activity of Cpl-1 v4 and Cpl-1 v5 was relatively low, so I continued with selected Pal variants: v3 and v9. I conducted an *ex vivo* assay by blocking their activity with WT Pal-specific serum, thus mimicking *in vivo* conditions where previous treatment with the WT endolysin induced considerable concentration of WT-specific IgG. I did not observe any decrease of Pal v3 activity when mixed with WT Pal-specific serum (comparing to activity in naïve serum). This observation demonstrates that Pal v3 variant escapes cross-reaction and resulting cross-neutralization by specific antibodies induced by WT Pal. Notably, the antibacterial activity of Pal v3 in the WT Pal-specific serum was significantly stronger than of WT Pal in the same conditions (Figure 20). I hypothesize, that two features of Pal v3 contributed to this advantage: it was not neutralized by serum specific to Pal and its basic bacteriolytic activity was higher than Pal. These data support the potential of Pal v3 in repeated use where induction of antibodies (by the first round of treatment with WT Pal) may be a concern.

This study demonstrates a new efficient approach using EndoScan for identification of antigenic epitopes in proteins. These epitopes can be selectively modified, without losing antibacterial activity of the endolysins, to decrease immunoreactivity of the proteins. To some extent, summary immunogenicity of a protein can be decreased (as in Pal v9), or its cross-reactions with wild-type-induced antibodies can be minimized (as in Pal v3). Thus, EndoScan can be used for optimization of active enzymes designed for therapeutic applications, by modifying immunogenicity of a protein

without sacrificing its activity, with a relatively low number of tested proteins. Importantly, this approach is universal and can be applied in many active proteins, so its applicability reaches far beyond endolysins and has potential in the design of numerous biological drugs.

11. Conclusions

Most crucial conclusion of this work in my opinion is that presented process characterized by wet-lab analysis, bioinformatics protein modelling and a next round of limited experiments yields modified enzymes that both retain (or even improve) their activity while showing modified biochemical feature (here: recognition of a protein by wild-type protein specific antibodies). Thorough testing in a first round of wet-lab experiments allows for precise protein modelling during bioinformatic analysis. It in turn allows to downsize enormous and impossible to calculate number of possible variants to literally a few variants with very high chance of success without loosing most important feature of the modified protein. This approach require precise planning of experiments during the first round, as data are crucial for eliminating variants with low or medium chance for success. Most variants tested herein show altered antibacterial activity, even though amino acid substitutions were not within catalytic site of the enzymes. This shows the importance of precisely determine what features cause a researched property of a protein. In return, bioinformatic analysis can isolate variants that shift as little features outside of the feature of interest as possible.

12. Summary

In presented work I investigated the potential for improving endolysin performance *in vivo*. I determined pre-clinical safety assessment of Cpl-1 and Pal endolysins. They induce a normal profile of specific IgG production, and no health threatening effects were noted. I developed a new method based on the use of DNA dye to precisely measure bacteriolytic activity of endolysins and their variants, including *ex vivo* conditions in biological samples. This method focuses on killing efficiency of an bacteriolytic agent. I also provided mathematical description of results that allows for direct comparison of activity for both endolysins and their variants in various environments or concentrations. I further identified amino acids that interact with the endolysin-specific IgG's in immunogenic regions of CPl-1 and Pal. I used this knowledge to propose potentially deimmunized variants of Cpl-1 and Pal. I designed and produced eight variants that retained antibacterial activity. All these variants were relatively immunogenic, but at the same time recognized weaker by wild type endolysins-specific antibodies, five of them demonstrated markedly decreased reactivity. All variants showed bacteriolytic activity: in five of them this activity was weaker than that of wild type endolysins, two retained the same activity, and one (Pal v3) showed increased bacteriolytic activity in comparison to wild type endolysin. Two variants of Pal endolysin (Pal v3 and pal v9) demonstrated no decrease in bacteriolytic activity when treated wild-type specific murine serum, while wild type Pal showed decrease in its enzymatic activity in the same conditions. Thus, I propose Pal v3 as the new highly active variant of endolysin Pal with improved performance *in vivo*, particularly less prone to neutralization by specific antibodies without losing bacteriolytic activity.

13. Streszczenie

W mojej pracy badałem możliwość poprawy działania endolizyn Cpl-1 i Pal w warunkach *in vivo*. Stwierdziłem, że obie te endolizyny indukują normalny, oczekiwany profil ekspresji przeciwciał specyficznych klasy IgG, nie zaobserwowałem żadnych innych oddziaływań potencjalnie negatywnych dla ludzkiego zdrowia. W toku swojej pracy opracowałem nową metodę wykorzystującą barwniki fluorescencyjne DNA do badania bakteriolitycznej aktywności endolizyn i ich wariantów, także w warunkach jakie występują w próbach biologicznych badanych *ex vivo*. Ta metoda skupia się na określaniu zdolności białka do niszczenia całych bakterii. Wraz z metodą dostarczyłem także matematyczną charakterystykę wyników pozwalającą na bezpośrednie porównanie aktywności obu endolizyn i ich wariantów w różnych warunkach lub stężeniach. Następnie w białkach Cpl-1 i Pal określiłem aminokwasy wchodzące w interakcję ze specyficznymi do nich przeciwciałami IgG. Na podstawie tej wiedzy zaproponowałem warianty tych endolizyn o potencjalnie obniżonej reaktywności immunologicznej. Wszystkie z tych wariantów okazały się wzbudzać niemal normalny poziom przeciwciał IgG *in vivo*, ale jednocześnie były rozpoznawane znacznie słabiej niż przeciwciała, które wzbudziła dzika forma endolizyn, pięć z tych wariantów było rozpoznawane znacząco słabiej. Wszystkie z zaproponowanych przeze mnie wariantów wykazywały aktywność bakteriolityczną, pięć z nich niższą, dwa z nich podobną, a jeden (Pal v3) wyższą w porównaniu do naturalnych endolizyn. Co więcej, dwa warianty endolizyny Pal (Pal v3 i Pal v9) wykazały brak wrażliwości na neutralizację przez przeciwciała specyficzne do dzikiej formy Pal i zachowały pełną aktywność bakteriolityczną.

Dlatego jako rezultat mojej pracy proponuję wariant Pal v3 jako nową endolizynę o szczególnym potencjale terapeutycznym, posiadającą zwiększoną aktywność bakteriolityczną, a jednocześnie obniżoną wrażliwość na neutralizację przez specyficzne przeciwciała indukowane przez dziką formę Pal.

14. Bibliography

1. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. How many species are there on earth and in the ocean? PLoS Biol. 2011;9(8):1–8.

2. Nodelman U, Allen C, Zalta EN. Stanford encyclopedia of philosophy. 2002;380.

3. Novikoff B. The concept of integrative levels and biology. Science (80-). 1945;101(2618):209–15.

4. Council NR. Synthesis and Manufacturing: Creating and Exploiting New Substances and New Transformations. In: Beyond the molecular frontier: challenges for chemistry and chemical engineering. 2003.

5. Palomares, LA, Estrada-Moncada S, Ramírez O. Recombinant gene expression. In: Production of recombinant proteins. 2004.

6. Palermo A, Chem D, Frsc E. Future of the Chemical Sciences. Chem Int. 2016;38(6).

7. Ponomarenko EA, Poverennaya E V., Ilgisonis E V., Pyatnitskiy MA, Kopylov AT, Zgoda VG, et al. The Size of the Human Proteome: The Width and Depth. Int J Anal Chem. 2016;2016.

8. Salingaros N. Some remarks on the algebra of Eddington's E numbers. Found Phys. 1985;15(6):683–91.

9. Eddington A, Newman J. The constants of nature. In: World of Mathematics. New York: Simon&Schuster; 1956.

10. Eddington A. The Fundamental Theory. 1946.

11. Eddington A. The philosophy of physical science: Tarner Lectures. Tarner Lect. 1938;

12. Carrascoza F, Zaric S, Silaghi-Dumitrescu R. Computational study of protein secondary structure elements: Ramachandran plots revisited. J Mol Graph Model [Internet]. 2014;50:125–33. Available from: http://dx.doi.org/10.1016/j.jmgm.2014.04.001

13. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts. 2010;1(3–4):271–83.

14. Zhang B, Li J, Lü Q. Prediction of 8-state protein secondary structures by a novel deep learning architecture. BMC Bioinformatics. 2018;19(1):1–13.

15. Kabsh W, Sander C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. Biophysics (Oxf). 1983;22:2577–637.

16. Jabeen A, Mohamedali A, Ranganathan S. Protocol for protein structure modelling. Encycl Bioinforma Comput Biol ABC Bioinforma. 2018;1–3(September 2017):252–72.

17. Richardson J. The Anatomy and Taxonomy of Protein Structure. Vol. 34, The American Journal of the Medical Sciences. 1981. 167–339 p.

18. Tompa P, Davey NE, Gibson TJ, Babu MM. A Million peptide motifs for the molecular biologist. Mol Cell [Internet]. 2014;55(2):161–9. Available from: http://dx.doi.org/10.1016/j.molcel.2014.05.032

19. Otaki JM. Availability of short amino acid sequences in proteins. Protein Sci. 2005;14(3):617–25.

20. Wardah W, Khan MGM, Sharma A, Rashid MA. Protein secondary structure prediction using neural networks and deep learning: A review. Comput Biol Chem [Internet]. 2019;81(August):1–8. Available from: https://doi.org/10.1016/j.compbiolchem.2019.107093

21. Gao M, Skolnick J. Structural space of protein-protein interfaces is degenerate, close to complete, and highly connected. Proc Natl Acad Sci U S A. 2010;107(52):22517–22.

22. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074–82.

23. Usmani SS, Bedi G, Samuel JS, Singh S, Kalra S, Kumar P, et al. THPdb: Database of FDA-approved peptide and protein therapeutics. PLoS One. 2017;12(7):1–12.

24. Vázquez R, García E, García P. Phage lysins for fighting bacterial respiratory infections: A new generation of antimicrobials. Front Immunol. 2018;9(OCT).

25. Abdelrahman F, Easwaran M, Daramola OI, Ragab S, Lynch S, Oduselu TJ, et al. Phage-encoded endolysins. Antibiotics. 2021;10(2):1–31.

26. Nelson D, Loomis L, Fischetti VA. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proc Natl Acad Sci U S A. 2001;98(7):4107–12.

27. (WHO) WHO. 2019 Antibacterial Agents in Clinical Development. 2019.

28. Fernández-Ruiz I, Coutinho FH, Rodriguez-Valera F. Thousands of novel endolysins discovered in uncultured phage genomes. Front Microbiol. 2018;9(MAY):1–8.

29. Rosselló-Móra R, Amann R. Past and future species definitions for Bacteria and Archaea. Syst Appl Microbiol [Internet]. 2015;38(4):209–16. Available from: http://dx.doi.org/10.1016/j.syapm.2015.02.001

30. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576–9.

31. Vollmer W, Blanot D, De Pedro MA. Peptidoglycan structure and architecture. FEMS Microbiol Rev. 2008;32(2):149–67.

32. Schneewind O, Missiakas D. Lipoteichoic acids, phosphate-containing polymers in the envelope of gram-positive bacteria. J Bacteriol. 2014;196(6):1133–42.

33. Buist G, Steen A, Kok J, Kuipers OP. LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol. 2008;68(4):838–47.

34. Yang H, Linden SB, Wang J, Yu J, Nelson DC, Wei H. A chimeolysin with extended-spectrum streptococcal host range found by an induced lysis-based rapid screening method. Sci Rep. 2015;5:17257.

35. Fischetti VA. Bacteriophage lysins as effective antibacterials. Curr Opin Microbiol. 2008;11(5):393–400.

36. Oliveira H, Melo LDR, Santos SB, Nobrega FL, Ferreira EC, Cerca N, et al. Molecular Aspects and Comparative Genomics of Bacteriophage Endolysins. J Virol. 2013;87(8):4558–70.

37. Hermoso JA, García JL, García P. Taking aim on bacterial pathogens: from phage therapy to enzybiotics. Curr Opin Microbiol. 2007;10(5):461–72.

38. P G, R L, C R, E G. Mechanism Of Phage-Induced Lysis In Pneumococci. Microbiol Soc. 1983;129(2).

39. Hermoso JA, Monterroso B, Albert A, Galán B, Ahrazem O, García P, et al. Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. Structure. 2003;11(10):1239–49.

40. Rau A, Hogg T, Marquardt R, Hilgenfeld R. A new lysozyme fold. Crystal structure of the muramidase from Streptomyces coelicolor at 1.65 Å resolution. J Biol Chem. 2001;276(34):31994–9.

41. Tews I, Terwisscha Van Scheltinga AC, Perrakis A, Wilson KS, Dijkstra BW. Substrate-assisted catalysis unifies two families of chitinolytic enzymes. J Am Chem Soc. 1997;119(34):7954–9.

42. Harhala M, Gembara K, Miernikiewicz P, Owczarek B, Kaźmierczak Z, Majewska J, et al. DNA Dye Sytox Green in Detection of Bacteriolytic Activity: High Speed, Precision and Sensitivity Demonstrated With Endolysins. Front Microbiol. 2021;12(October):1–10.

43. Lunjani N, O'Mahony L. Critical entry points-microbiome. Implement Precis Med Best Pract Chronic Airw Dis. 2018;15–9.

44. Dasgupta S, Erturk-Hasdemir D, Ochoa-Reparaz J, Reinecker HC, Kasper DL. Plasmacytoid dendritic cells mediate anti-inflammatory responses to a gut commensal molecule via both innate and adaptive mechanisms. Cell Host Microbe [Internet]. 2014;15(4):413–23. Available from: http://dx.doi.org/10.1016/j.chom.2014.03.006

45. World Health Organization. Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. Geneva; 2020.

46. Nikkaido H. Multidrug Resistance in Bacteria. Annu Rev Biochem. 2009;78:119–46.

47. (WHO) WHO. Global Action Plan on Antimicrobial Resistance. Vol. 10, World Health Organisation. 2015.

48. (WHO) WHO, (FAO) F and AO of the UN, (OIE) WO for AH. Antimicrobial Resistance: A Manual for Developing National Action Plans. Jama. 2016.

49. WHO. Global Report on Surveillance 2014. WHO 2014 AMR Rep [Internet]. 2014;1–8. Available from: http://www.who.int/drugresistance/documents/AMR_report_Web_slide_set.pdf

50. (WHO) WHO. WHO global strategy for containment of antimicrobial resistance, World Health Organisatin [Internet]. World Health Organisation. 2001. Available from: http://apps.who.int/iris/handle/10665/66860

51. Serra-Burriel M, Keys M, Campillo-Artero C, Agodi A, Barchitta M, Gikas A, et al. Impact of multi-drug resistant bacteria on economic and clinical outcomes of healthcare-associated infections in adults: Systematic review and meta-analysis. PLoS One. 2020;15(1):1–14.

52. Gierke; Ryan, Wodi; Patricia, Miwako K. Pneumococcal disease. Pract Nurse. 2015;45(10):255–74.

53. Loh B, Leptihn S. A Call For a Multidisciplinary Future of Phage Therapy to Combat Multi-drug Resistant Bacterial Infections. Infect Microbes Dis. 2020;2(1):1–2.

54. Rohde C, Wittmann J, Kutter E. Bacteriophages: A therapy concept against multi-drug-resistant bacteria. Surg Infect (Larchmt). 2018;19(8):737–44.

55. Labrou N. Therapeutic Enzymes: Function and Clinical Implications. Vol. 1148. Springer; 2019.

56. Gembara K, Dąbrowska K. Phage-specific antibodies. Curr Opin Biotechnol. 2021;68:186–92.

57. Sugathan S, Pradeep NS, Abdulhameed S. Bioresources and bioprocess in biotechnology. Bioresour Bioprocess Biotechnol. 2017;2:1–438.

58. Alcalde M. Directed Enzyme Evolution: Advances and Applications. Directed Enzyme Evolution: Advances and Applications. 2017. 1–284 p.

59. du Sert NP, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, et al. Reporting animal research: Explanation and elaboration for the arrive guidelines 2.0. Vol. 18, PLoS Biology. 2020. 1–65 p.

60. Harhala M, Nelson DC, Miernikiewicz P, Heselpoth RD, Brzezicka B, Majewska J, et al. Safety studies of pneumococcal endolysins Cpl-1 and Pal. Viruses. 2018;10(11).

61. Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, et al. Comprehensive serological profiling of human populations using a synthetic human virome. Science (80-). 2015;348(6239):1–23.

62. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

63. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: An online force field. Nucleic Acids Res. 2005;33(SUPPL. 2):382–8.

64. Pérez-Dorado I, Campillo NE, Monterroso B, Hesek D, Lee M, Páez JA, et al. Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. J Biol Chem. 2007;282(34):24990–9.

65. Ambrish Roy, Alper Kucukural, Yang Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2005;5(4):45–67.

66. Lewis CL, Craig CC, Senecal AG. Mass and density measurements of live and dead gram-negative and gram-positive bacterial populations. Appl Environ Microbiol. 2014;80(12):3622–31.

67. McKellar RC, Knight K. A combined discrete-continuous model describing the lag phase of Listeria monocytogenes. Int J Food Microbiol. 2000;54(3):171–80.

68. Yu ACS, Loo JFC, Yu S, Kong SK, Chan TF. Monitoring bacterial growth using tunable resistive pulse sensing with a pore-based technique. Appl Microbiol Biotechnol. 2014;98(2):855–62.

69. Meyers A, Furtmann C, Jose J. Direct optical density determination of bacterial cultures in microplates for high-throughput screening applications. Enzyme Microb Technol [Internet]. 2018;118(April):1–5. Available from: https://doi.org/10.1016/j.enzmictec.2018.06.016

70. WHO. 2020 Antibacterial Agents in Clinical and Preclinical Development. World Health Organization 2021. Geneva, Switzerland; 2021. 76 p.

71. Jun SY, Jang IJ, Yoon S, Jang K, Yu KS, Cho JY, et al. Pharmacokinetics and Tolerance of the Phage endolysin-based candidate drug SAL200 after a single intravenous administration among healthy volunteers. Antimicrob Agents Chemother. 2017;61(6):1–11.

72. Entenza JM, Loeffler JM, Grandgirard D, Fischetti VA, Moreillon P. Therapeutic effects of bacteriophage Cpl-1 lysin against Streptococcus pneumoniae endocarditis in rats. Antimicrob Agents Chemother. 2005;49(11):4789–92.

73. Witzenrath M, Schmeck B, Doehn JM, Tschernig T, Zahlten J, Loeffler JM, et al. Systemic use of the endolysin Cpl-1 rescues mice with fatal pneumococcal pneumonia. Crit Care Med. 2009;37(2):642–9.

74. Jado I, López R, García E, Fenoll A, Casal J, García P, et al. Phage lytic enzymes as therapy for antibiotic-resistant Streptococcus pneumoniae infection in a murine sepsis model. J Antimicrob Chemother. 2003;52(6):967–73.

75. Ainsworth S. Steady-State Enzyme Kinetics. Steady-State Enzyme Kinetics. 1977.

76. Lebaron P, Catala P, Parthuisot N. Effectiveness of SYTOX green stain for bacterial viability assessment. Appl Environ Microbiol. 1998;64(7):2697–700.

77. Mitchell GJ, Nelson DC, Weitz JS. Quantifying enzymatic lysis: Estimating the combined effects of chemistry, physiology and physics. Phys Biol. 2010;7(4).

78. Tafoya DA, Hildenbrand ZL, Herrera N, Molugu SK, Mesyanzhinov V V., Miroshnikov KA, et al. Enzymatic characterization of a lysin encoded by bacteriophage EL. Bacteriophage. 2013;3(2):e25449.

79. Bucevičius J, Lukinavičius G, Gerasimaite R. The use of hoechst dyes for DNA staining and beyond. Chemosensors. 2018;6(2).

80. Okamoto A. Next-generation fluorescent nucleic acids probes for microscopic analysis of intracellular nucleic acids. Appl Microsc. 2019;49(1):1–7.

81. Quyen TL, Ngo TA, Bang DD, Madsen M, Wolff A. Classification of Multiple DNA Dyes Based on Inhibition Effects on Real-Time Loop-Mediated Isothermal Amplification (LAMP): Prospect for Point of Care Setting. Front Microbiol. 2019;10(October):1–12.

82. (WHO) WHO. Global Framework for Development & Stewardship to Combat Antimicrobial Resistance [Internet]. Https://Www.Who.Int/Antimicrobial-Resistance/En/. 2018. Available from: https://www.who.int/phi/news/WHO_OIE_FAO_UNEP_Working_paper_of_the_framework_FINAL.pdf

83. United States Government. National Action Plan For Combating Antibiotic-resistant Bacteria. 2015.