SH3 BINDING DOMAINS FROM PHAGE ENDOLYSINS: HOW TO USE THEM FOR DETECTION OF GRAM-POSITIVE PATHOGENS

Veronika Jarábková^{1*}, Lenka Tišáková¹, Martin Benešík², Andrej Godány¹

Address(es): Veronika Jarábková,

¹University of Ss. Cyril and Methodius in Trnava, Faculty of Natural Sciences, Department of Biology, Nám. J. Herdu 2, 917 01 Trnava, Slovakia, (+421) 33 / 55 65 321.

²Masaryk University, Faculty of Science, Department of Experimental Biology, Kotlářská 2, Brno 611 37, Czech Republic.

*Corresponding author: veronika.jarabkova@gmail.com

doi: 10.15414/jmbfs.2020.9.6.1215-1220

ARTICLE INFO

Received 20. 5. 2019
Revised 11. 3. 2020
Accepted 17. 3. 2020
Published 1. 6. 2020

Regular article

ABSTRACT

Phage-encoded enzymes, called endolysins, could be linked to the purpose of specifically detecting and treating Gram-positive pathogens, such as drug-resistant bacteria *Staphylococcus aureus* or *Streptococcus agalactiae*. Most endolysins encoded by staphylococcal phages are composed of a C-terminal SH3b type domain, an amidase catalytic domain in the central position and an N-terminal catalytic domain with endopeptidase activity. Furthermore, endolysins from a streptococcal phage, e.g. B30 consist of a dual-enzyme active domain (EAD)- muramidase and endopeptidase and one or several cell wall binding domains (CBD), in this case SH3. An understanding of CBDs and domain design that has a high degree of host specificity may lead to shifts in the lytic spectrum of a chimeric endolysin. Our research was focused on the *in silico* analysis of CBDs of endolysins encoded by staphylococcal and streptococcal phages. Gene for specific binding domain (SBD) was designed based on bioinformatics approaches. The coding gene sequence of the SBD domain was subsequently fused with the gene for GFP (green fluorescent protein) and subcloned into the pET21 expression vector. The domain was prepared as a fusion recombinant protein and analyzed for its binding function. This capability was shown by the fusion of SBD with fluorescent tag – GFP, followed by fluorescent microscopy. Binding assays showed that SBD was able to bind to living cells of *Staphylococcus aureus* and *Streptococcus agalactiae*. Because of these properties, CBDs may be used as tools in alternative diagnostic methods. The fusions of SBD and EAD may bring the chimeric endolysin for being utilized as novel therapeutic agents.

Keywords: phage, endolysin, SH3 binding domain, *Staphylococcus aureus*, *Streptococcus agalactiae*

INTRODUCTION

Endolysins are phage-encoded lytic enzymes with the ability to disrupt cell walls that present possible alternatives to conventional antibiotics. Endolysins are characterized by two functions – specificity and enzymatic activity – allowed by their functional domains. The elimination of bacteria is only one of many possible applications for endolysins. Many studies have revealed that endolysins are capable of binding to the cell wall (Huang *et al.*, 2015; Yang *et al.*, 2017). This understanding has a lot of potential to be applied medically, such as in sensitive and rapid detection of bacteria for food quality and clinical safety (Yu *et al.*, 2015). Endolysins of Gram-positive infecting phages usually use modular structure (Díaz *et al.*, 1990) that is formed by enzymatically active domains (EADs) and cell wall binding domain (CBDs) (Figure 1). CBDs allow the phage to bind non-covalently to the unique carbohydrate components of bacterial peptidoglycan and may be responsible for the specificity of strain- or species binding. CBDs can recognize and bind to specific receptors on the bacterial cell wall and thereby increase the affinity of binding proteins. Flexible peptide linkers connect the domains in the construction of chimeric enzymes with new features. The modular organization of endolysins gives a unique opportunity for protein engineering, enabling the modification of protein activity, solubility, specificity and other physicochemical features (Kashani *et al.*, 2017).

Staphylococcus aureus and *Streptococcus agalactiae* belong to a group of foodborne pathogens in products (e.g. milk), which are contaminated during the handling of raw ingredients (Yu *et al.*, 2016). In the light of the above-mentioned problems, the detection of these bacteria is of foremost importance for food safety and quality. Nowadays, traditional methods for the detection of bacterial contamination, such as cultivation, are still time-consuming and laborious (Zanardi *et al.*, 2014). Due to this, alternative detection methods have been very desired and welcomed. It is assumed that the specificity of CBDs from phage endolysins can be utilized for the rapid detection of foodborne pathogens. CBDs with different specificity were developed to allow the detection and

differentiation of *Staphylococcus aureus* (Dong *et al.*, 2015). One of the earliest identified sequences of CBDs was the lysostaphin SH3b domain (Src homology 3). SH3 domains are encoded and expressed by various organisms (e.g. bacteria, yeasts, humans) and involved in different cellular processes, such as endocytosis (Gkourtsa *et al.*, 2016).

Figure 1 Typical modular organisation of endolysin from phage infecting Gram-positive bacteria. N-terminus with enzymatically active domain, C-terminus with binding domain, L- linker (adapted from Jarábková *et al.* 2015).

These domains mediate protein-protein interaction, which are important for linking intracellular signaling pathways and regulating the catalytic activity of proteins. The evolutionary conservation of SH3 domains provides evidence for their added value, such as importance for specific species targeting, e.g. for endopeptidase domain of lysostaphin and Ale-1 to bind to *S. aureus* (Becker *et al.*, 2009). However, some experimental data provides conflicting evidence, which does not always support the essential role of SH3b in endolysins. It is known that some SH3 domains require an intact pentaglycine bridge for binding function (Kashani *et al.*, 2017), which was demonstrated for SH3 domain of the bacteriocin lysostaphin (Grundling *et al.*, 2006).

Protein engineering highlights the potential of chimeric endolysins with a particular application. A CBD – GFP (green fluorescent protein) fusion constructs allow effective non-destructive host monitoring (Schmelcher *et al.*, 2011; Tišáková *et al.*, 2013; Chang *et al.*, 2017). GFP is a universal biological

marker for monitoring physiological processes, visualizing protein localization, and detecting transgenic expression *in vivo* (Loessner et al., 2002). Numerous derivatives of fluorescent proteins have been developed according to the color spectrum, e.g. green, blue, and yellow (Tsien, 1998; Shaner et al., 2007).

Catalytic domains - cysteine, histidine-dependent amidohydrolase / peptidase (CHAP) domain and glycosyl hydrolase family 25, CP family of lysozymes (GH25_Cpl-1-like), with located glycohydrolase (Glyco_hydro_25 domain), have been described in endolysins from phage infecting *St. agalactiae* (Oliveira et al., 2013). Additionally, most endolysins derived from staphylococcal phages contain SH3b domain (Becker et al., 2009; Kashani et al., 2017).

Here we describe *in silico* analysis and rational design, as well as expression, purification and subsequent binding assay of an engineered CBD = "SBD" - the specific binding domain of endolysin, which can recognize and bind to staphylococcal and streptococcal cells – *S. aureus* and *St. agalactiae*, specifically. In addition, we evaluate and observe the binding function of SBD against bacterial strains *S. aureus* and *St. agalactiae*.

MATERIALS AND METHODS

Design of specific binding domain

Bioinformatics analysis was oriented on the polyvalent *Staphylococcus* and *Streptococcus* phage endolysins. Nucleotide and protein sequences of endolysins representing phage-encoded lytic proteins (available to February 2019) were taken from NCBI (<http://www.ncbi.nlm.nih.gov/>; Schütz et al., 2000). Information about identity was retrieved from the GenBank database (<http://www.ncbi.nlm.nih.gov/>). Subsequently, the identity of each sequence was confirmed in the Uniprot database (<http://www.uniprot.org/>; Apweiler et al., 2004). Conserved domains – CHAP, PGRP, SH3 - were identified in individual endolysins, and for the verification of protein sequences, CDD (Marchler-Bauer et al., 2008) and PFAM (Finn et al., 2014) were used. Protein and nucleotide sequences were also aligned using CLUSTALW2 (Larkin et al., 2007) on the EBI server (Rodríguez-Tomé et al., 1996). The resulting alignments were manually edited and colored in MS-WORD 2013, labelling either invariant or conserved residues as well.

Bioinformatics analysis of the cell wall binding domain from phage B30

The secondary structure of binding domain from endolysin phage B30 was predicted by using I-Tasser (<http://zhanglab.ccmb.med.umich.edu/I-Tasser/>) and SWISS-MODEL (Arnold et al., 2006). Prediction of tertiary structure models of SH3 binding domain from phage endolysin infecting *St. agalactiae* was created in PHYRE2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) (Kelley and Sternberg, 2009). For multiple structural alignments of protein structure, MultiProt server (<http://bioinfo3d.cs.tau.ac.il/MultiProt/>; Shatsky et al., 2004) was used. Superimpositions of tertiary structures were edited and visualized in WebLabViewerLite 4.0.

Rational design of SBD

SH3 binding domain of endolysin from phage B30 was selected as a template for the design of SBD. This domain was the only identified binding domain of endolysin from phage infecting *St. agalactiae*. Potentially removable parts of amino acids sequence of SH3 binding domain were found in the multiple sequence alignment of binding domains. These regions did not contain any conserved amino acids. Subsequently, secondary structure prediction of SH3 proved the presence of α -helices and β -sheets in the sequence. Tertiary structure confirmed the occurrence of amino acids, supposedly playing a role in binding function, in the regions with no conserved amino acids as discussed in results.

Experimental part

Based on *in silico* analyses, nucleotide sequence of specific binding domain (*sbd*) has been proposed - CBD was designed for specific targeting bacterial cell wall of selected bacteria – *Staphylococcus aureus* and *Streptococcus agalactiae*.

Bacterial strains, vectors, media and grown conditions

In this study, *Escherichia coli* MC1061 strain (Casadaban and Cohen, 1980) was used for cloning experiments. *Escherichia coli* BL21 (DE3) (Novagen, Germany) was used as a host for the production of recombinant protein SBD-GFP. Both *E. coli* strains were grown at 37 °C in Luria-Bertani medium - peptone (10 g/L), yeast extract (5 g/L), NaCl (10 g/L), distilled water (1000 mL), pH = 7.5 (Maniatis et al., 1982), after plasmid transformation into *E. coli* supplemented with the selection marker – ampicillin (50 µg/mL) (AppliChem, Germany), under shaking conditions at 120 rpm (Sambrook and Russel, 2001). Bacterial strains for binding assays were cultivated at 37 °C - *Staphylococcus aureus* in THB (Todd Hewitt Broth, Sigma-Aldrich, USA), for *Streptococcus agalactiae* in TSB (Tryptic Soy Broth, Sigma-Aldrich, USA).

Vector carrying synthetic binding domain - pUC19-sbd (Novagen, Germany), vectors pET-21 (Novagen, Germany) and pET21-gfp (Tišáková et al., 2014) were used for cloning and expression of the gene for specific binding domain (SBD). Construct pET21-sbd-gfp was prepared by conventional recombinant DNA techniques, as discussed below. Bacterial strains *Staphylococcus aureus* CCM 4028, *Streptococcus agalactiae* CCM 6187 used for the binding assay were obtained from Czech Collection of Microorganisms (<http://www.sci.muni.cz/ccm>).

Construction of recombinant plasmid construct

Constructions of recombinant vectors carrying genes for SBD (specific binding domain) protein (5.43 kDa) and GFP (green fluorescent protein) (26.89 kDa) were performed using expression vector pET21. For linearization of pET21, restriction enzymes (NcoI and HindIII, New England Biolabs, USA) were added, with corresponding Tango buffer (New England Biolabs, USA), with reaction set for 2 h at 37 °C, according to manufacturer's instructions. After linearization, pET21 was dephosphorylated by SAP (Shrimp Alkaline Phosphatase, Thermo Scientific, USA) and reaction was set to 1 h at 37 °C, followed by thermal inactivation for 20 min at 80 °C. Insert *sbd* (gene for specific binding domain) was excised with NcoI and HindIII (New England Biolabs, USA) from pUC19-sbd. Insert *gfp* (gene for green fluorescent protein) was excised with HindIII and XhoI (New England Biolabs, USA) from pET21-gfp. Inserts were purified by GeneJet Gel Extraction Kit (Qiagen, Germany), according to manufacturer's instructions. Vector pET21, insert *sbd* and insert *gfp* were ligated 1 h at 22 °C by Rapid DNA Ligation Kit (New England Biolabs, USA), followed by thermal inactivation at 65 °C for 20 min. Subsequently, ligation mixture was transformed into *E. coli* Top10F. The result of ligation was the pET21-sbd-gfp construct. Correct ligation was verified by DNA sequencing (Eurofins Genomic, Germany).

Engineered SBD-GFP protein production

E. coli BL21 (DE3) cells, containing recombinant plasmids (1 µL of recombinant plasmid was used for transformation), were grown overnight at 37 °C in LB medium supplemented with ampicillin (50 µg/mL). The culture was subcultured 1:100 into fresh medium, grown at 37 °C under shaking to the optical density OD_{600nm} ~ 0.5. Expression of gene (*sbd-gfp*) was induced by 0.4 mM isopropyl β -D-1-thiogalactopyranoside (IPTG, Sigma-Aldrich, USA), after which bacteria were shaken overnight at 18 °C. Cells were harvested from 1,000 mL cultures (6,000 rpm, 10 min, 4 °C). Lysis of cells was mediated by BugBuster MasterMix (Novagen, Germany) supplemented with ProteoBlock Protease Inhibitor Cocktail (Thermo Scientific, USA) and shaking at room temperature for 30 min. The soluble protein fraction of the cell homogenate was obtained by centrifugation (16,000 rpm, 15 min, 4 °C).

Engineered SBD-GFP protein isolation and analysis

Partially purified protein SBD-GFP (32 kDa) was obtained by IMAC (immobilized metal affinity chromatography). Nickel-nitrilotriacetic acid (Ni-NTA) Agarose (Qiagen, Germany) was added to a cleared supernatant and mixed softly at 4 °C for 40 min. Proteins were eluted from a column with elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, and 250 mM imidazole). Protein concentration was carried out using Amicon Ultra-4 Centrifugal Filter Units (10 kDa) (Merck Millipore, Germany). Protein was stored in Tris-HCl buffer (50 mM Tris-HCl, pH 8.0 + 1 mM dithiothreitol). Samples of eluates were then assayed for purity via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein purity was assessed on 12 % gel, while Spectra Multicolor Broad Range Ladder (Thermo Scientific, USA) was used as a molecular size marker. Visualization of proteins on gel was supported by Coomassie brilliant blue G250 (Sigma-Aldrich, USA) with staining procedure for 15 min at room temperature and destaining via conventional methods. Finally, recombinant protein was stored in Tris-HCl buffer (50 mM Tris-HCl, pH 8.0 + 1 mM dithiothreitol) at -80 °C until binding assays were performed. The final concentration of the recombinant protein was determined via a NanoDrop spectrophotometer.

Engineered SBD-GFP binding function assays

As a standard assay to determine the binding ability of SBD-GFP towards bacterial surface of *S. aureus* and *St. agalactiae* cells – the conditions from Tišáková et al. (2013) were adopted with some modifications. Briefly, strains were cultivated 18 h, 120 rpm and harvested by centrifugation, resuspended in 1:10 volume of Tris-HCl (50 mM, pH 8.0), 40 µL bacterial suspension and 20 µL of diluted fusion protein SBD-GFP, or GFP as negative control, standing at room temperature for 10 min. Subsequently, cells were harvested by centrifugation (5 min, 11,000 rpm) and supernatant was discarded, cell pellets were washed twice in Tris-HCl buffer and finally resuspended in 15 µL of selected buffer. Cells were visualized by microscopy in bright field, as well as with fluorescence, in order to investigate GFP fluorescence (1000x) all images were acquired using a LEICA DM2500 microscope equipped with a LEICA DFC290 HD camera and LAS software.

RESULTS AND DISCUSSION

Bioinformatics analysis of staphylococcal and streptococcal phage endolysins

In this study, all identified staphylococcal and streptococcal phage endolysins containing CBDs, especially SH3 type, were compared.

Multiple sequence alignments of nucleotide and protein sequences

Comparisons of amino acids and nucleotide sequences (Figure 2) binding domains of endolysins from staphylococcal and streptococcal bacteriophages were obtained. Based on the data from NCBI database, seven endolysins from polyvalent staphylococcal phages and one endolysin from streptococcal bacteriophage B30 (Tab 1) were selected for further analyses. It must be noted that in our database results, CBD has been identified as a functional domain specifically in endolysin from streptococcal phage B30. Other endolysins from phages attacking *St. agalactiae* strain do not display an identified CBD of the SH3 family which was the main focus of this study. Seven invariant amino acid residues and 22 corresponding positions of amino acid residues are shown in amino acid comparisons, which signify the different level of amino acid conservation in endolysin binding domains sequences from different phages (Zhang et al., 2012). There are two amino acid residues, W52 and Y58, whose numbering is based on their order in the protein sequence, that are considered conserved. The potential binding function of these two amino acid residues was supported by a tertiary structure of SH3 binding domain of endolysin from B30 in which these two amino acids were among the five potential binding sites. Due to this, they could likely affect the binding function in the entire binding domain.

Table 1 Selected endolysin from polyvalent staphylococcal and streptococcal phages.

n	phage	strain	References	ID		Domains		
				Uniprot	GenBank	CHAP	PGRP	SH3
1	P4W	<i>S. aureus</i>	Łobocka <i>et al.</i> , 2012	I6XK14	AFN38929	+	+	+
2	MSA6	<i>S. aureus</i>	Łobocka <i>et al.</i> , 2012	I6WKT7	AFN38718	+	+	+
3	GH15	<i>S. aureus</i>	Gu <i>et al.</i> , 2012	D6QY02	ADG26756	+	+	+
4	FI200W	<i>S. aureus</i>	Łobocka <i>et al.</i> , 2012	I6X5A8	AFN38507	+	+	+
5	A3R	<i>S. aureus</i>	Łobocka <i>et al.</i> , 2012	I6W8P7	AFN38100	+	+	-
6	phi 812	<i>S. aureus</i>	Pantůček <i>et al.</i> , 1998	A1YTR6	ABL87142	+	+	+
7	Mutant 812F1	<i>S. aureus</i>	Benešik <i>et al.</i> , 2018	A1YTR6	KJ206562	+	-	+
8	B30	<i>St. agalactiae</i>	Pritchard <i>et al.</i> , 2004	Q8HA43	AAN28166	+	GH25_Cpl-1-like Glyco_hydro_25	+

Prediction of secondary structure of SH3 binding domain Our hypothesis about the possible deletion of the mentioned part (30 amino acids) from the sequence of the proposed synthetic binding domain of interest was verified by secondary structure prediction analysis of the selected CBD of endolysin from phage B30. The α -helices and β -sheets in the structure of the proposed sequence were obtained. Five to six α -helices were demonstrated to be present in the K12-K16 and A39-D44 regions. In these regions, it was believed to be inappropriate to divide or delete the sequence, as most stable conformation (α -helix) of the lowest energy polypeptide chain could be formed here (Elliott et al., 1997). In the α -helix structure, the most common amino acids are alanine, glutamic acid (E13, E43), leucine and methionine. Among the α -helix-rich regions, there are two regions – R20-K26 and W30-R32, with the presumed occurrence of β -pleated sheet. The formation of β -pleated sheet is supported by amino acids such as valine (V21, V23), isoleucine and phenylalanine. To broaden our knowledge, the predicted secondary structure of SH3 domain of endolysin from B30 were compared, acquired by the Lomets tool, with predicted secondary structure of the same sequence by SWISS-MODEL server. Predictions were comparable in terms of the α -helix and the β -pleated sheet content. In the SWISS-MODEL predictions, the most stable areas with 2 α -helix (2 α -helices) are: S9-K16 and S38-V42. This stands in correspondence with Lomets predictions, where these regions are shifted in the first case by three amino acids and in the second region by one amino acid. In the case of β -pleated sheet, the match of the results is even higher. In the R20-V27 region, the shift is only in one amino acid. Differences that occur in the predictions of the secondary structure are uniquely related to the various algorithms, which are used by these tools for predictions.

Tertiary structure prediction of SH3 binding domain

The prediction of tertiary structure was performed using the structure recognition method on the online PHYRE server. In this method of 3D models design, which belongs to the inductive group, the knowledge of the known patterns is used by this algorithm. Tertiary structures were designed according to templates selected by an algorithm. Each result is indicated with a corresponding number (ID) in PDB. For the following brief analysis, the two models that were chosen reflected

Figure 2 The aligned amino acid binding domains sequences of endolysins from phages, namely P4W, MSA6, GH15, FI200W, A3R, phi812 infecting *S. aureus* and endolysin from phage B30 infecting *St. agalactiae*. Alignment was prepared in Clustal Omega at the EBI server. Invariant amino acid residues are coloured by blue and signed with symbol (*); corresponding substitutions are coloured by grey and signed with (:) for conservative substitutions and (.) for semi-conservative substitutions.

Experimental demonstration of the binding functions of W52 and Y58 occurring in the SH3 binding domains of endolysins expressed by bacteria infecting *S. aureus* and *St. agalactiae* has still not been recognized yet. There are not any amino acid residues, that are considered as conserved, from the G21 to the W52 (Figure 2). This fact led to the idea that any residue from the G21 to the W52 cannot be deleted. Despite the highest degree of conservation of SH3 domain among the different groups of enzymes (Kashani et al., 2017), our bioinformatics results showed that the length of SH3 domain from selected staphylococcal endolysins is 67 amino acids and SH3 domain of endolysin from streptococcal phage B30 is 48 amino acids. It is worth noting that the modular organization of SH3b-containing endolysins is highly similar, as shown in Table 1, but the divergence is considerable at the amino acid level (Becker et al., 2009), which was confirmed by our alignments.

the highest value of confidence and identity. In the first case, the template named 2KYB is a mannose-glycoprotein molecule endo- β -n-acetylglucosaminidase from *Clostridium perfringens*. In the second case, the template 2KSR was assigned to SH3 from the same bacterium, *C. perfringens*. Attention was paid to tertiary structures using the MultiProt server (<http://bioinfo3d.cs.tau.ac.il/MultiProt/16.2.2019>) (Figure 3). Tertiary structures of the SH3 domain of B30 phage endolysin obtained in PHYRE were compared with tertiary structures of templates that had been determined by program as initial for model designing. Some amino acids were identified as potential binding sites in the 3D model of binding domain, which was obtained from Lomets program on I-Tasser.

Figure 3 Structural alignment of the SH3 domain of endolysin from B30 phage (red line) and template (blue line). Structures were obtained in the PHYRE and translated at the MultiProt. Highlighted amino acid residues determines potential binding sites, views were created in WebLabViewerLite 4.0. Translated models of structures were shown with ribbon type, atom – off and amino acids residues were characterized by balls and sticks.

These binding sites are recognized by specific binding domains and play a role in the interaction between the binding domain and its PG substrate. Five amino acids were marked and compared in the structural alignment that represented potential binding sites (Figure 4). As it can be seen from the figures, the structures are very similar and differ only in parts where the overlap is not complete. Amino acids designated as potential binding sites do not correspond in all cases. In the first case - template 2KYB - the pairs of amino acids E47 - N34 and F43 - W30 were not matched. In the second case - template 2KRS - differences were observed in pairs of amino acid residues M1 - V1 and R46 - N34 (Figure 3). The occurrence of amino acids is displaced by several amino acid sites, for the reason that the templates were represented by longer structures compared in respect to the length of SH3.

Figure 4 Tertiary structure model of SH3 binding domain of endolysin from bacteriophage B30 infecting *St. agalactiae* with predicted potential binding sites - active amino acid residues (obtained by Lomets on I-Tasser server). Purple digits and letters indicate predicted potential binding sites.

Rational design of specific binding domain

As a template for the design of SBD, the CBD of endolysin from phage B30 infecting *St. agalactiae* was determined. As discussed below, the intention of the design was to prepare CBD with the ability to bind to *S. aureus* and *St. agalactiae* cells. It should be stressed, that rational design was supported by bioinformatics analysis for the aspects of protein folding, maintaining of binding function and potentially binding amino acids in sequence. As seen from multiple sequence alignments, some regions do not contain any conserved amino acid. In the event of potential exclusion of a particular segment of a sequence, protein secondary structure was also taken into account, whether there is a functional α -helix or β -sheet in the region, so that after deletion of the selected sequence, the sequence forming the secondary structure is not disrupted. Kong et al. (2017) described an engineered CBD for the detection of *Bacillus cereus*, where the idea of rapid, sensitive, eco-friendly development of diagnostic methods of bacteria was supported. The designed amino acid sequence and nucleotide sequence with start codon for synthetic gene coding specific binding domain is described below:

Amino acid sequence

MVESKPNASSADKEFIKAGTRVRVYEKVNWSRINHPESAQWVEDNY

Nucleotide sequence

ATGGTAGAGTCTAAGCCAAACGCAAGTAGCGTGATAAAGAATTTAT
CAAGGCAGGAAGCTCGTGTAAAGAGTTTATGAAAAAGTGAATGGATGGT
CACGCATTAACCATCCAGAGTCGGCGCAATGGGTAGAAGATAACTAC

Engineered SBD-GFP

GFP as a part of our fusion protein SBD-GFP, serves to visualize the binding function of this recombinant protein. Purified protein SBD-GFP was used in binding assays to evaluate the ability of the recombinant protein to bind to *Staphylococcus aureus* and *Streptococcus agalactiae* cells (Figure 5). GFP was used as a negative control, with no observed binding function.

Supposedly, the varying length of protein sequence of CBDs means variable protein folding depends on the primary structure and amino acid sequence and GFP which forms larger part of the fusion proteins could play a significant role. This presumption should be empirically verified in more protein-oriented studies. Another point of future studies would be to precisely identify the part of CBD which is exactly responsible for full binding function. It was demonstrated that SBD binds to the *S. aureus* and *St. agalactiae* by the binding assay. This fusion protein is expected to be used in the preparation of chimeras with other types of binding domains to detect and visualize various Gram-positive bacterial species

or with catalytic domains, such as CHAP domain to inhibit Gram-positive bacteria or in protein mixture.

Figure 5 Confocal fluorescent microscopy - the binding function of fusion protein SBD-GFP to the cell surface of *S. aureus* and *St. agalactiae* and the single GFP as a negative control. A) Fluorescent microscopy B) Transmitted light microscopy 1) *St. agalactiae* CCM 6187 2) *S. aureus* CCM 4028 3) negative control; magnification 1000x.

Frequently, SH3 domain occurs in the modular structure of endolysins from phages infecting *Staphylococcus*, but it can be possibly found in other endolysins, such as endolysins from phage infecting *Streptococcus*. This unique CBD may be exploited for the enhancement of anti-staphylococcal efficacy (e.g. in protein fusion constructs) (Becker et al., 2009; Jarábková et al., 2015). The SH3 domain of the lysostaphin was the one of the first identified binding domains discovered, mentioned CBDs prefers to bind to the proline rich sequence (Mayer and Eck, 1995). The new efficient and cost-effective methods for detection assay are nowadays essential in the development of diagnostic tools. The use of antibodies in therapeutics has gained its popularity, however, the production of soluble antibodies is time consuming and very expensive (Kong et al., 2017).

In many studies, the fusion of CBDs and fluorescent markers, such as GFP, were used for binding assays (Loessner et al., 2002; Dong et al., 2014; Tišáková et al., 2014). SH3b binding domain of endolysin LysGH15B was one prime example of chimeric endolysin where binding function was proved. In the fusion protein LysGH15B-GFP, fluorescent tag - GFP allowed the visualization of specific affinity to staphylococcal host, especially the strains resistant to methicillin (MRSA = methicillin-resistant *Staphylococcus aureus*). Fusion protein GFP-LysGH15B bound successfully to the 43 strains of *S. aureus* and *S. epidermidis* (Gu et al., 2011; Benešik et al., 2017). Another significant report of an engineered endolysin with an increased lytic activity is Ply187AN-KSH3b, a fusion of the endopeptidase from staphylococcal endolysin Ply187 with SH3b domain from another staphylococcal endolysin LysK. The fusion construct was more than 10-fold more active in multiple activity assay than original endolysins (Mao et al., 2013).

Differences in the peptidoglycan layer among these bacteria were taken into account – distinctions are usually present on the level of glycan strand and peptide bridge, while the size of interpeptide bridge is one or seven amino acids. On the other hand, the occurrence of amino acids is different: L-alanine, L- or D-serine is typically in peptidoglycan layer of *S. aureus* (Mainardi et al., 2008). In the peptidoglycan layer of streptococcal bacteria from group B, where also *St. agalactiae* is classified, the interpeptide bridge is formed by L-alanine-L-alanine or L-alanine-L-serine among L-lysine on the one side of the peptide strand and D-alanine on the other side of peptide strand.

In previous studies, fusion proteins were used for endolysin analysis focused on binding function of CBDs. Majority of works have been devoted to peptidoglycan hydrolases from *Clostridium* and *Pneumococcus* and in heterologous fusions, due to the ability of endolysins to bind to their host cells (Díaz et al., 1991; Croux et al., 1993; Loessner et al., 2002; Mayer et al., 2011; Mao et al., 2013). CBDs are responsible for the specific activity of many endolysins, thanks to their ability to bind specific ligands in the cell walls (e.g. peptidoglycan subunits, carbohydrates, proteins, lipoteichoic acid) (Loessner, 2005). Ligands recognized by binding domains of endolysins have not been sufficiently identified, but presumed specific ligands are predominantly cell wall carbohydrates of which recognition and binding function of a binding domain to a susceptible host has been confirmed (Schmelcher et al., 2010, 2012). In Gram-positive bacteria, at least eight different types of peptidoglycan are known (Schleifer and Kandler, 1972). *Staphylococcus* strains possess type A3 α with penta-glycin structure (Kamisanog et al., 1982; Schumann, 2011). Subgroup A3 is cross-linked by interpeptide bridges consisting of either monocarboxylic L-amino acids or glycine or sometimes both. *Streptococcus* species also belong to type A3 α , but the size of the interpeptide bridge is between 2 to 7 amino acid residues (Schleifer and Kandler, 1972). It could be predicted that the similarity of cell wall composition might be ultimate for the binding function of SBD to staphylococcal and streptococcal cells. Additionally, pentaglycine bridge might be attached by SBD because it is already known as an important location for binding of SH3 of lysostaphin (Grundling and Schneewind 2006). CBDs are proposed for binding to cell wall, fusions between CBDs and GFP demonstrated their highly specific recognition ability (Loessner, 2005). Generally, fluorescent protein markers are used in molecular biology, medicine, but also in other disciplines, due to its convenient properties like stability and non-toxicity (safety) (Schmelcher et al., 2010). Here it was used to demonstrate the binding function of SBD to live bacterial cells. Fluorescence microscopy was utilized for the confirmation of binding function to the staphylococcal and streptococcal cultures. Use of this method is applicable in development of new detection methods of bacteria and also in modular engineering – for instance, protein with a high affinity to the bacterial cell wall (Kretzer et al., 2007; Schmelcher et al., 2010). In conclusion, CBDs tagged with a fluorescence protein represent a very promising tool for developing an alternative method of detection for specific bacterial species. The binding function is expected to be very specific, therefore the occurrence of false positive results should be very low.

CONCLUSIONS

The bioinformatics analyses in this study comprised all endolysins encoded by phages infecting *Staphylococcus aureus* and *Streptococcus agalactiae* with characterized CBDs available in the public database. These analyses supported the rational design of SBD with binding function to *S. aureus* and *St. agalactiae*. The practical focus of this study was the preparation of purified protein SBD-GFP followed by binding experiments with substrate from living *S. aureus* cells, respectively. *St. agalactiae*. The binding function of SBD was analyzed by fluorescence microscopy, using GFP fusion proteins. Additionally, the designed specific binding domain was successful at targeting selected bacterial cells.

Acknowledgements: The present work was funded by APVV-16-0173; FPPV-14-2017; FPPV-12-2018; M.B. was supported by grant MUNI/A/0958/2018 from the Grant Agency of the Masaryk University. Part of experimental work was done in Department of Experimental Biology at Masaryk University in Brno. We wish to thank Roman Pantůček (Masaryk University) for supporting in the part of experimental work and Markéta Pernisová (CEITEC, Masaryk University) for the fluorescent microscopy.

REFERENCES

APWEILER, R., BAIROCH, A., WU, C.H., BARKER, W.C., BOECKMANN, B., FERRO, S., GASTEIGER, E., HUANG, H., LOPEZ, R., MAGRANE, M., MARTIN, M.J., NATALE, D.A., O'DONOVAN, C., REDASCHI, N., YEH, L.S.L. 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*, 32, D115-D119. <http://dx.doi:10.1093/nar/gkh131>

ARNOLD, K., BORDOLI, L., KOPP, J., SCHWEDE, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2), 195-201. <http://dx.doi:10.1093/bioinformatics/bti770>

BECKER, S. C., FOSTER-FREY, J., STODOLA, A. J., ANACKER, D., DONOVAN, D. M. 2009. Differentially conserved staphylococcal SH3b_5 cell

wall binding domains confer increased staphylolytic and streptolytic activity to a streptococcal prophage endolysin domain. *Gene*, 443(1-2), 32 – 41. <http://dx.doi:10.1016/j.gene.2009.04.023>

BENEŠÍK, M., NOVÁČEK, J., JANDA, L., DOPITOVÁ, R., PERNISOVÁ, M., MELKOVÁ, K., TIŠÁKOVÁ, L., DOŠKAŘ, J., ŽÍDEK, L., HEJÁTKO, J., PANTUČEK, R. 2017. Role of SH3b binding domain in a natural deletion mutant of *Kayvirus* endolysin LysF1 with a broad range of lytic activity. *Virus Genes*, 54(1), 130-139. <http://dx.doi:10.1007/s11262-017-1507-2>

CASADABAN, M.J., COHEN, S.N. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of Molecular Biology*, 138, 179–207.

DÍAZ, E., LOPEZ, R., GARCIA, J. L. 1990. Chimeric phage-bacterial enzymes: A clue to the modular evolution of genes. *Proceedings of the National Academy of Sciences of the United States of America*, 87(20), 8125 – 8129.

CROUX, C., RONDA, C., LÓPEZ, R., GARCÍA, J. L. 1993. Interchange of functional domains switches enzyme specificity: construction of a chimeric pneumococcal-clostridial cell wall lytic enzyme. *Molecular Microbiology*, 9(5), 1019-25.

DONG, Q., WANG, J., YANG, H., WEI, C., YU, J., ZHANG, Y. M., HUANG, Y., ZHANG, X. E., WEI, H. 2014. Construction of chimeric lysin Ply187N-V12C with extended lytic activity against staphylococci and streptococci. *Microbe Biotechnology*, 8(2), 210-20. <http://dx.doi:10.1111/1751-7915.12166>

FINN, R.D., TATE, J., MISTRY, J., COGGILL, P.C., SAMMUT, S.J., HOTZ, H.R., CERIS, G., FORSLUND, K., EDDY, S.R., SONNHAMMER, E.L.L., BATEMAN, A. 2014. The Pfam protein families database. *Nucleic Acids Research*, 36, 281-288. <http://dx.doi:10.1093/nar/gkm960>

GKOURTSA, A., BURG, J., AVULA, T., HOCHSTENBACH, F., DISTEL, B. 2016. Binding of a proline-independent hydrophobic motif by the *Candida albicans* Rvs167-3 SH3 domain. *Microbiological Research*, 190, 27-36. <http://dx.doi.org/10.1016/j.micres.2016.04.018>

GRUNDLING, A., SCHNEEWIND, O. 2006. Cross-Linked Peptidoglycan Mediates Lysostaphin Binding to the Cell Wall Envelope of *Staphylococcus aureus*. *Journal of Bacteriology*, 188, 2463-2472. <http://dx.doi:10.1128/jb.188.7.2463-2472.2006>

GU, J., LU, R., LIU, X. 2011. LysGH15B, the SH3b domain of staphylococcal phage endolysin LysGH15, retains high affinity to staphylococci. *Current Microbiology*, 63(6), 538-542. <http://dx.doi:10.1007/s00284-011-0018-y>

CHANG, Y., RYU, S. 2017. Characterization of novel cell wall binding domain-containing *Staphylococcus aureus* endolysin LysSA97. *Applied Microbiology and Biotechnology*, 101(1), 147-158. <http://dx.doi:10.1007/s00253-016-7747-6>

HUANG, Y., YANG, H., YU, J., WEI, H. 2015. Molecular dissection of phage lysin PlySs2: integrity of the catalytic and cell wall binding domains is essential for its broad lytic activity. *Virologica Sinica*, 30(1), 45 – 51. <http://dx.doi:10.1007/s12250-014-3535-6>

JARÁBKOVÁ, V., TIŠÁKOVÁ, L., GODÁNY, A. 2015. Phage endolysin: a way to a binding function of C-terminal domains a mini review. *Nova Biotechnologica et Chemica*, 14(2), 117-134. <http://dx.doi:10.1515/nbec-2015-0021>

KAMISANGO, K., SAIKI, I., TANIO, Y., OKUMURA, H., ARAKI, Y., SEKIKAWA, I., AZUMA, I., YAMAMURA, Y. 1982. Structures and biological activities of peptidoglycans of *Listeria monocytogenes* and *Propionibacterium acnes*. *Journal of Biochemistry*, 92, 23–33.

KASHANI, H. H., SCHMELCHER, M., SABZALIPOOR, H., HOSSEINI, E. S., MONIRI, R. 2017. Recombinant endolysins as potential therapeutics against antibiotic-resistant *staphylococcus aureus*: current status of research and novel delivery strategies. *Clinical Microbiology Reviews*, 31(1), e00071-17. <http://dx.doi:10.1128/cmr.00071-17>

KELLEY, L. A., STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*, 4(3), 363-371. <http://dx.doi:10.1038/nprot.2009.2>

KRETZER, J.W., LEHMANN, R., SCHMELCHER, M., BANZ, M., KIM, K., KORN, C., LOESSNER, M. J. 2007. Use of high-affinity cell wall-binding domains of bacteriophage endolysins for immobilization and separation of bacterial cells. *Applied and Environmental Microbiology*, 73(6), 1992 – 2000. <http://dx.doi:10.1128/aem.02402-06>

KONG, M., SHIN, H.J., SUNGGI, H., PARK, J.K., RYU, S. 2017. Lateral flow assay-based bacterial detection using engineered cell wall binding domain of a phage endolysin. *Biosensors and Bioelectronics*, 96, 173-177. <http://dx.doi:10.1016/j.bios.2017.05.010>

LARKIN, M.A., BLACKSHIELDS, G., BROWN, N.P., CHENNA, R., MCGETTIGAN, P.A., MCWILLIAM, H., VALENTIN, F., WALLACE, I.M., WILM, A., LOPEZ, R., THOMPSON, J.D., GIBSON, T.J., HIGGINS, D.G. 2007. Clustal Wand Clustal X version 2.0. *BMC Bioinformatics*, 23(21), 2947–2948. <http://dx.doi:10.1093/bioinformatics/btm404>

LOBOCKA, M., HEJNOWICZ, M.S., DABROWSKI, K., GOZDEK, A., KOSAKOWSKI, J., WITKOWSKA, M., ULATOWSKA, M.I., WEBER-DABROWSKA, B., KWIATEK, M., PARASION, S., GAWOR, J., KOSOWSKA, H., GLOWACKA, A. 2012. Genomics of staphylococcal Twtort-like phages–potential therapeutics of the post-antibiotic era. *Advances in Virus Research*, 83, 143–216.

<http://dx.doi.org/10.1016/B978-0-12-394438-2.00005-0>

- LOESSNER, M. J., KRAMER, K., EBEL, F., SCHERER, S. 2002. C-terminal domains of *Listeria monocytogenes* bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. *Molecular Microbiology*, 44(2), 335 – 49. <http://dx.doi.org/10.1046/j.1365-2958.2002.02889.x>
- LOESSNER, M.J., REES, C.E.D. 2005. *Listeria* phages: basics and applications. Waldor, M.K., Friedmann, D.I., Adhya, S.L., Washington DC, 362-379 p., ISBN 1-55581-307-0.
- LOW, L.Y., YANG, C., PEREGO, M., OSTERMAN, A., LIDDINGTON, R.C. 2005. Structure and lytic activity of a *Bacillus anthracis* prophage endolysin. *The Journal of Biological Chemistry*, 280 (42), 35433-35439. <http://dx.doi.org/10.1074/jbc.M502732000>
- MARCLER-BAUER, A., ANDERSON, J.B., CHITSAZ, F., DERBYSHIRE, M.K., DE-WEESE-SCOTT C., FONG, J.H., GEER, L.Y., GEER, R.C., GONZALES, N.R., GWADZ, M., HE, S., HURWITZ D.I., JACKSON, J.D., KE, Z., LANCZYCKI, CH.J., LIEBERT, C.YA., LIU, CH., LU, F., LU, S., MARCLER, G., MULLOKANDOV, M., SONG, J.S., TASNEEM, A., THANKI, N., YAMASHITA, R.A., ZHANG, D., ZHANG, N., BRYANT, S.H. 2008. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Research*, 37(1), D205-D210. <http://dx.doi.org/10.1093/nar/gkn845>
- MAO, J., SCHMELCHER, M., HARTY, W. J., FOSTER-FREY, J., DONOVAN, D.M. 2013. Chimeric Ply187 endolysin kills *Staphylococcus aureus* more effectively than the parental enzyme. *FEMS Microbiology Letters*, 342, 2013, 30-36. <http://dx.doi.org/10.1093/nar/gkn845>
- MANIATIS, T., FRITSCH, E.F., SAMBROOK, J.: Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1982, 545 p., ISBN 0-8796-9136-0.
- MAYER, B. J., ECK, M. J. 1995. SH3 domains. *Current Biology*, 5(4),364-7.
- PANTŮČEK, R., ROSYPALOVÁ, A., DOŠKAŘ, J., KAILEROVÁ, J., RUŽIČKOVÁ, V., BORECKÁ, P., SNOPOKOVÁ, S., HORVÁTH, R., GOTZ, F., ROSYPAL, S. 1998. The polyvalent staphylococcal phage phi 812: its host-range mutants and related phages. *Virology*, 246(2), 241–252.
- OLIVEIRA, H., MELO, L.D.R., SANTOW, S.B., NÓBREGA, F.L., FERREIRA, E.C., CERCA, N., AZEREDO, J., KLUSKENS, L.D. 2013. Molecular aspect and comparative genomics of bacteriophage endolysins. *Journal of virology*, 87(8), 4558-4570. <http://dx.doi.org/10.1128/JVI.03277-12>.
- PRITCHARD, D. G., DONG, S., BAKER, J. R., ENGLER, J. A. 2004. The bifunctional peptidoglycan lysin of *Streptococcus agalactiae* bacteriophage B30. *Microbiology*, 150 (7), 2079-87. <http://dx.doi.org/10.1099/mic.0.27063-0>
- REGULSKI, K., COURTIN, P., KULAKAUSKAS, S., CHAPOT-CHARTIER, M. 2013. A novel type of peptidoglycan-binding domain highly specific for amidated D-Asp cross-bridge, identified in *Lactobacillus casei* bacteriophage endolysins. *Journal of Biological Chemistry*, 288(8), 20416-20426. <http://dx.doi.org/10.1074/jbc.M112.446344>
- RODRIGUEZ-TOME, P., STOEHR, P.J., CAMERON, G.N., FLORES, T.P. 1996. The European Bioinformatics Institute (EBI) database. *Nucleic Acids Research*, 24(1), 6-12.
- SAMBROOK, J., RUSSEL D.W. 2001. *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor, Laboratory Press, Cold Spring Harbor, NY, 2001. ISBN 10 0-87969-577-3.
- SHANER, N. C., PATTERSON, G. H., DAVIDSON, M. W. 2007. Advances in fluorescent protein technology. *Journal of Cell Science.*, 120, 4247-60. <http://dx.doi.org/10.1242/jcs.005801>
- SHATSKY, M., NUSSINOV, R. WOLFSON, H.J.: A method for simultaneous alignment of multiple protein structures. *Proteins*, 56(1), 143-156. <http://dx.doi.org/10.1002/prot.10628>
- SCHLEIFER, K.H., KANDLER, O. 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Clinical Microbiology Reviews*, 407-477.
- SCHMELCHER, M., SHABAROVA, T., EUGSTER, M. R. 2010. Rapid multiplex detection and differentiation of *Listeria* cells by use of fluorescent phage endolysin cell wall binding domains. *Applied and Environmental Microbiology*, 76(17), 5745 - 5756. <http://dx.doi.org/10.1128/AEM.00801-10>
- SCHMELCHER, M., WALDHER, F., LOESSNER, M. J. 2011. *Listeria* bacteriophage peptidoglycan hydrolases feature high thermoresistance and reveal increased activity after divalent metal cation substitution. *Applied Microbiology and Biotechnology*, 93(2), 633 -643. <http://dx.doi.org/10.1007/s00253-011-3372-6>
- SCHUMANN, P. 2011. Peptidoglycan Structure. London: Academic Press. 101-129 p.. Edited by F. Rainey & A. Oren. <http://dx.doi.org/10.1016/B978-0-12-387730-7.00005-X>
- SCHÜTZ, J., COPLEY, R.R., DOERKS, T., PONTING, C.P., BORK, P. 2000. Smart: a webbased tool for the study of genetically mobile domains. *Nucleic Acids Research*. 28231– 234.
- TIŠÁKOVÁ, L., VIDOVA, B., FARKAŠOVSKÁ, J., GODÁNY, A. 2013. Bacteriophage endolysin 1/6: characterization of the C-terminal binding domain. *Microbiology Letters*, 350, 199-208. <http://dx.doi.org/10.1111/1574-6968.12338>
- TIŠÁKOVÁ, L., GODÁNY, A. 2014. Bacteriophage endolysins and their use in biotechnological processes. *The Journal of Microbiology, Biotechnology and Food Sciences*, 350, 164-170.
- TSIEN R. Y. 1998. The green fluorescent protein. *Annual Review of Biochemistry*, 67, 509-44.
- YANG, H., ZHANG, H., WANG, J., YU, J., WEI, H. 2017. A novel chimeric lysin with robust antibacterial activity against planktonic and biofilm methicillin-resistant *Staphylococcus aureus*. *Scientific Reports*, 7, 40182. <http://dx.doi.org/10.1038/srep40182>
- YU, J., ZHANG, Y., ZHANG, Y., LI, H., YANG, H., WEI, H. Sensitive and rapid detection of *Staphylococcus aureus* in milk via cell binding domain of lysin. *Biosensors and Bioelectronics*, 77, 366-71. <http://dx.doi.org/10.1016/j.bios.2015.09.058>
- ZANARDI, G., CAMINITI, A., DELLE DONNE, G., MORONI, P., SANTI, A., GALLETI, G., TAMBA, M., BOLZONI, G., BERTOCCHI, L. 2014. Short communication: Comparing real-time PCR and bacteriological cultures for *Streptococcus agalactiae* and *Staphylococcus aureus* in bulk- tank milk samples. *Journal of Dairy Science*, 97(9), 5592-8. <http://dx.doi.org/10.3168/jds.2014-7947>.
- ZHANG, X. 2012. Diversity analysis of peptidoglycan hydrolases in lactic acid bacteria and their phages. *International Dairy Journal*, 25, 60-65. <http://dx.doi.org/10.1016/j.idairyj.2012.02.003>