

SEARCHING AND IN SILICO CHARACTERIZATION OF STREPTOMYCES PHAGE ENDOLYSINS AND THEIR CATALYTIC DOMAINS

Darina Peterková*¹, Zuzana Šramková^{1,2}, Michaela Mancoš¹, Andrej Godány¹

Address(es):

¹University of Ss. Cyril and Methodius in Trnava, Faculty of Natural Sciences, Department of Biology Nám. J. Herdu 2, 91701 Trnava, Slovakia, +421903878862.

²Slovak Academy of Sciences, Institute of Molecular Biology, Dubravska cesta 21, 845 51 Bratislava, Slovakia.

*Corresponding author: darinapeterkova@gmail.com

doi: 10.15414/jmbfs.2020.10.2.221-229

ARTICLE INFO

Received 12. 5. 2020
Revised 15. 6. 2020
Accepted 18. 6. 2020
Published 1. 10. 2020

Regular article



ABSTRACT

Endolysins, peptidoglycan hydrolases encoded by bacteriophages, degrade bacterial cell wall and are promising alternative to antibiotics. However, the selection and bioengineering of endolysins needs previous bioinformatic characterization. This study focuses on endolysins encoded by viruses infecting *Streptomyces* spp., describes *Streptomyces* phages diversity and association. Set of 143 endolysins were predicted in 175 *Streptomyces* phage genomes from NCBI database and The Actinobacteriophage Database and were characterized their phylogeny and function domains, especially catalytic domains. Predicted endolysins have amidase (Ami_2 and CHAP domain), muramidase (glycol_hydro_25) and peptidase activity (CHAP, NlpC/P60). This bioinformatic characterization serve as a base to next research in developing of endolysins with new properties in enzybiotik form.

Keywords: *Streptomyces* spp., endolysin, *in silico*, bioinformatic, function domains

INTRODUCTION

Phages or bacteriophages are specialized viruses able to infect host species belonging to the domain Bacteria. Importance of phages in the treatment of bacterial infections has been recognized in Eastern Europe several decades ago (Voelker, 2019). However, due to an increasing problem with multi-drug resistant bacteria, Western medicine has also begun to pay attention to this approach in recent years. The traditional phage therapy is associated with some risks such as adaptation of viable viruses to target other, closely related strains, transduction of virulence genes, lysogenic conversion of commensal bacteria and bacterial resistance to phages (Valero-Rello, 2019). For this reason, not the whole phage particles but products of their genes, lytic enzymes such as endolysins, represent more useful tool against multiresistant bacteria. Endolysins are also attractive for other fields, for example, in biotechnology for primary industry (Hoopes et al., 2009; Schmelcher et al., 2015); in food industry as the prevention against contamination by foodborne bacterial pathogens (Schmelcher and Loessner, 2016); and as a diagnostic tool to detect bacteria in different types of samples (Kretzer et al., 2007; Bai et al., 2016; Gómez-Torres et al., 2018).

Endolysins are bacteriophage-encoded enzymes which cleave peptidoglycan (PG) of their host bacterial cell wall at the end of phage lytic cycle (Young, 1992). This degradation results in osmotic shock, cell rupture and often bacteria death (Fischetti, 2008). The lytic mechanism is effective also after exogenous application of purified endolysins (Nelson et al., 2001; Loeffler et al., 2001; Cheng et al., 2007; Schmelcher et al., 2012). The specificity of endolysins (from genus to strain specificity) makes endolysins a potential alternative or additive to current antibiotics as enzybiotics (Schmelcher et al., 2012; Nelson et al., 2001). Depending on the reaction and type of bond in PG being cleaved by certain endolysin, these enzymes can be categorized into several classes: amidases, glucosaminidases, transglycosylases, lysozymes and endopeptidases (Oliveira et al., 2012). Amidases hydrolyze the amide bond connecting MurNAC to the peptide stems; glucosaminidases, transglycosidases and lysozyme cleave the glycosidic bonds in glycan chain and endopeptidases act on the peptide bonds forming the peptide bridge or stem of PG.

Generally, endolysins of phages infecting Gram-positive bacteria are modular proteins composed of N-terminally located enzymatic active domain (EAD) and C-terminally located cell wall binding domain (CBD). The domains are connected by short flexible linkers (Fenton et al., 2010; Schmelcher et al., 2012). The EAD is responsible for cleaving a specific bond in PG and usually there is only one EAD present in endolysin structure. The cell wall binding domain is responsible for recognizing specific epitopes on the bacterial cell wall

and there can be one or more CBDs present in one endolysin (Schmelcher et al., 2012; Oliveira et al., 2012).

Lot of endolysins were predicted, but the number of biochemically confirmed endolysins is lower and only few structures are known, especially endolysins infecting Gram-positive bacteria.

Our research is devoted to Streptomyces phage endolysins because *Streptomyces* spp. are interesting for their high antibiotic production, since 70 % of clinically used antibiotics are from this species and on proteomic level, this genus is similar to *Mycobacterium* spp., such as *Mycobacterium tuberculosis* (Smith et al., 2013). This work is focused on searching for new endolysins encoded by *Streptomyces* phages. *In silico* analysis performed in this study aims to phage genome mining, identification of genes with endolysin function and especially to characteristics of catalytic domains of these peptidoglycan degrading enzymes. All the information represents a basis for the future use of these enzymes in developing new types of chimeric endolysins and enzybiotics.

MATERIALS AND METHODS

Phage genomes

Whole-genome sequences of phages infecting *Streptomyces* spp. were obtained from the Actinobacteriophage Database (<https://phagesdb.org/>) and NCBI (<https://www.ncbi.nlm.nih.gov/>) Genome and Nucleotide databases. Information regarding host species, phage life cycle, taxonomy, and continent where the host streptomyces was isolated, was also collected from these databases. If the data were not accessible, they were searched for in other sources, publications and databases.

Identification of endolysins and their functional domains

Open reading frames were predicted by annotation tools PHASTER (<http://phaster.ca/>; Arndt et al., 2016; Zhou et al., 2011) and BASys (<https://www.basys.ca/>; Van Domselaar et al., 2005). If neither of these tools predicted gene for endolysin in phage genomes, Protein BLAST with blastp algorithm (protein-protein BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used for functional prediction of gene products and searching for the endolysin. For function domain was used putative conserved domains prediction in graphic summary of blastp which is linked to CD-search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>; Marchler-Bauer et al., 2017). CD-search also predicted amino acid residues in catalytic site, which were marked in CLC Sequence viewer.

Multiple sequence alignment

For identification of conserved amino acid residues, whole endolysin sequences were aligned by ClustalW (Jeannot et al., 1998) integrated in CLC Sequence viewer 8.0. Conserved amino acid residues were identified in aligned sequences of the catalytic domains extracted from whole endolysin sequences. Graphic representations of multiple sequence alignments were performed using WebLogo (weblogo.berkeley.edu; Crooks et al., 2004) by creating sequence logos for each functional domain type identified in the studied set of endolysin sequences. Conserved amino acid residues predicted in catalytic site were marked in each WebLogo.

Phylogenetic trees

The neighbour-joining algorithm was used to calculate the phylogenetic trees from whole endolysins sequence alignment and alignments of catalytic domains (Saitou and Nei, 1987), applying the Jones-Taylor-Thornton model (Jones et al., 1992) of the amino acid changes, both integrated in CLC Sequence viewer 8.0. Interactive Tree Of Life (iTOL) (Letunic and Bork, 2019) was used to visualize phylogenetic trees.

Protein tertiary structure models

On 3D level, models were created by Phyre2 *ab initio* modeling (Kelley et al., 2015). 3D models were visualized by PyMOL(TM) 2.0.6, in this program were also made structure alignments and visualization of conserved amino acid residues in active site. To compute the protein charge Protein isoelectric point calculator (<http://isoelectric.org/calculate.php>) was used at pH 7.4.

RESULTS

Streptomyces phages

The set of 175 fully sequenced Streptomyces phage genomes was obtained in this study. These genomes were divided into 18 groups according to the host from which the phages were isolated (Table 1). *Streptomyces griseus* was found to represent the most common host (n=65, 38%), followed by *S. lividans* (n=19, 11.10%), *S. xanthochromogenes* (n=18, 10.5%) and *S. venezuelae* (n=16, 9.4%). Other host *Streptomyces* are represented at a level below 5%: *S. platensis* and *S. scabiei* (n=8, 4.7%); *S. coelicolor* (n=6, 3.5%); *S. azureus* (n=5, 2.9%); *S.*

viridochromogenes, *S. avermitilis*, *S. toxytricini*, *S. griseofuscus* and *Streptomyces sp.* (n=4, 2.3%); *S. flavovirens* (n=2, 1.2%) and *S. aureofaciens*, *S. roseosporus*, *S. virginiae*, *S. himastatinicus* (n=1, 0.6%) (Table 1). Almost all phages of the set belong to the order Caudovirales represented by the families *Siphoviridae* (155), *Podoviridae* (10) and *Myoviridae* (1). Only one member of the studied set of phages was classified to the family *Tectiviridae* and 4 phages were unclassified. No visible connection between phage family and host was observed, most of the isolated hosts are infected by phage belonging to different families. Except *S. scabiei* and *S. spp.*, these hosts are infected by phages belonging to three families. Most of the *Streptomyces* phages are lysogenic (n=81, 56.6%), but none of them are purely lysogenic, and about a third of phages from the set has lytic life cycle only (n=48, 33.6%). For the rest of the phages (n=14, 9.8%) data were not found. Most of the phages were isolated from North America (n=148, 84.6%).

Endolysins and their domains

Endolysins, proteins involved in late host lysis, were predicted in 143 genomes from 175 studied genomes and classified into 5 groups according the activity: amidases, muramidases, peptidases, cell wall binding and others (associated with glycosidehydrolases but with unclear function). Most of the endolysins were predicted with modular domain structure (Figure 2, Table 2), where on the N-terminus is EAD (Ami_2, CHAP, NLPC/P60, gglyco_hydro_25) and on C-terminal end is CWBD (LysM, PG_binding_1, CW_7). Fn3 domain was included because this protein is presented in the location near the gene for holine, where endolysin is usually located, its function is still not clear (Valk et al., 2017) and this domain is connected with glycosidehydrolases. Generally, streptomyces actinophage endolysins have the length about 300 amino acids (226 AA phiCAM – 471 AA NooNoot and Paradiddles). Most of endolysins contain one catalytic and one cell wall binding domain (50) and numerous contain one additional cell wall binding domain (11). In cases where only one domain was predicted, there has been usually space for missing domain (catalytic or cell wall binding, Figure 2). Domains repeats are related with fn3 domain and CW_7, which occur with two and three repeats on the C-terminus, respectively. In phylogenetic tree of *Siphoviridae* virus endolysins (Figure 3A) are endolysins with CHAP or NLPC/P60 domain grouped together and on for endolysins from phages *Podoviridae* family have not been predicted any EAD, only CWBD PG_binding_1 domain. Endolysins from this family all have a negative net charge.

Table 1 General characterization of studied Streptomyces phages

Total phages	GC%	Family					Lytic	Lysogeny	Unknown data	Continent				Data not found	
		Siphoviridae	Podoviridae	Tectiviridae	Myoviridae	Unclassified				Africa	Asia	Europe	North America		
Ssp	4	65.1	1 (25)	0	0	1 (25)	2 (50)	0	2 (50)	2 (50)	0	1 (25)	0	1 (25)	2 (50)
Shi	1	70	1 (100)	0	0	0	0	0	1 (100)	0	0	0	0	0	1 (100)
Sgf	4	54.1	4 (100)	0	0	0	0	4 (100)	0	0	0	0	0	4 (100)	0
Svg	1	59.1	1 (100)	0	0	0	0	1 (100)	0	0	0	0	0	1 (100)	0
Sfl	2	69.1	2 (100)	0	0	0	0	0	0	2 (100)	2 (100)	0	0	0	0
Sro	1	70.3	0	0	0	0	1 (100)	0	1 (100)	0	0	0	0	0	1 (100)
Sau	1	71.2	1 (100)	0	0	0	0	0	0	1 (100)	0	0	1 (100)	0	0
Sto	4	65.4	4 (100)	0	0	0	0	1 (25)	3 (75)	0	0	0	0	4 (100)	0
Sav	4	66.6	3 (75)	0	0	0	1 (25)	0	1 (25)	3 (75)	0	2 (50)	1 (25)	0	1 (25)
Svd	4	58.8	3 (75)	1 (25)	0	0	0	3 (75)	1 (25)	0	0	0	0	4 (100)	0
Sco	6	65.4	6 (100)	0	0	0	0	0	5 (83.3)	1 (16.7)	0	0	4 (66.7)	0	2 (33.3)
Ssc	8	59.4	6 (75)	1 (12.5)	1 (12.5)	0	0	7 (87.5)	0	1 (12.5)	0	2 (25)	0	6 (75)	0
Sxa	18	66.5	18 (100)	0	0	0	0	0	18 (100)	0	0	0	0	18 (100)	0
Sve	16	69.4	16 (100)	0	0	0	0	0	15 (93.8)	1 (6.2)	0	0	1 (6.2)	15 (93.8)	0
Spl	8	66.6	8 (100)	0	0	0	0	1 (12.5)	7 (87.5)	0	0	0	0	8 (100)	0
Saz	5	64.7	5 (100)	0	0	0	0	1 (20)	3 (60)	1 (20)	0	0	0	5 (100)	0
Sli	19	56.2	18 (94.7)	1 (5.3)	0	0	0	7 (36.8)	9 (47.4)	3 (15.8)	0	0	1 (5.3)	18 (94.7)	0
Sgi	65	60.1	58 (89.2)	7 (10.8)	0	0	0	49 (75.4)	16 (24.6)	0	0	1 (1.5)	0	64 (98.5)	0

Sgr – *S. griseus*; Sli – *S. lividans*; Saz – *S. azureus*; Spl – *S. platensis*; Sve – *S. venezuelae*; Sxa – *S. xanthochromogenes*; Ssc – *S. scabiei*; Sco – *S. coelicolor*; Svd – *S. viridochromogenes*; Sav – *S. avermitilis*; Sto – *S. toxytricini*; Sau – *S. aureofaciens*; Sro – *S. roseosporus*; Sfl – *S. flavovirens*; Svg – *S. virginiae*; Sgf – *S. griseofuscus*; Shi – *S. himastatinicus*; Ssp – *Streptomyces spp.* The percentage is in parentheses.

Table 2 Predicted endolysins and their domains in *Streptomyces* phages

Ssp	3	2	1	0	0	1	1	0	0
Shi	1	0	0	1	0	0	0	0	0
Sgf	1	1	0	0	0	0	2	0	0
Svg	1	0	0	0	0	0	0	0	1
Sfl	2	2	0	0	0	0	2	0	0
Sro	1	0	1	0	0	0	1	0	0
Sau	1	0	0	0	0	1	0	0	0
Sto	4	2	0	0	0	3	2	0	0
Sav	4	1	1	0	0	2	0	0	0
Svd	3	2	0	0	0	2	1	0	0
Sco	6	2	1	0	2	2	0	0	0
Ssc	7	1	0	0	0	0	0	0	6
Sxa	17	10	1	0	3	14	6	0	0
Sve	16	11	0	0	0	12	7	0	0
Spl	8	3	3	0	1	2	0	1	0
Saz	4	0	2	0	0	3	1	0	1
Sli	16	6	3	0	2	8	4	0	1
Sgi	47	28	7	0	0	28	16	3	3

Activity	Predicted endolysins	N-acetylmuramoyl-L-alanine amidase (EC 3.5.128)	cysteine, histidine-dependent amidohydrolases/peptidases (CHAP) Amidase or peptidase	N-acetylmuramidase (lysozyme, EC 3.2.1.17)	Cysteine peptidase	Peptidoglycan binding	Though to anchor to extracellular polysaccharides such as peptidoglycan and chitin	Peptidoglycan binding	Not clear role in glycosidase hydrolases
Protein Family		Amidase_2	CHAP	Glyco_hydro_25	NLPC_P60	PG_binding_1	LysM	CW_7	fn3
Pfam domain		PF01510	PF05257	PF01183	PF00877	PF01471	PF01476	PF08230	PF00041
Category		Amidases	Muramidases	Peptidases	Cell wall binding	Other			

Sgr – *S. griseus*; Sli – *S. lividans*; Saz – *S. azureus*; Spl – *S. platensis*; Sve – *S. venezuelae*; Sxa – *S. xanthochromogenes*; Ssc – *S. scabiei*; Sco – *S. coelicolor*; Svd – *S. viridochromogenes*; Sav - *S. avermitilis*; Sto – *S. toxytricini*; Sau – *S. aureofaciens*; Sro – *S. roseosporus*; Sfl – *S. flavovirens*; Svg – *S. virginiae*; Sgf – *S. griseofuscus*; Shi – *S. himastatinicus*; Ssp – *Streptomyces spp*

Amidases

Amidases (EC 3.5.128) are the most presented group (n=91). Amidases have Zn²⁺ depending N-acetylmuramoyl-L-alanine amidase activity and cleave 1,6-anhydro-MurNac-L-Ala bonds in MurNac-tri and tetrapeptides. In *Streptomyces* phage the group of endolysins is presented by Amidase_2 (PF01510) and CHAP domain (PF05257). Most of studied endolysins were predicted with Amidase_2 or/and PGRP domain (Figure 3A), which are the same catalytic domains defined by different databases and by CD-search to belong to PGRP superfamily (Marchler-Bauer et al., 2017) and mostly contain as cell wall binding domains PG_binding_1 or/and LysM. domain. In this case the host or cycle specificity was not recorded, but almost all endolysins in this group are from endolysins belonging to *Siphoviridae* family and have positive net charge. According to the alignment and CD search domain prediction of all predicted Amidase_2 and PGRP domains, was shown two main groups. First group contains Zn²⁺ binding residues His6, His123 and Cys131, and catalytic residues Tyr47 and Thr19. His11, His149 and Asp159 are predicted to be zinc binding residues and Cys42, Lys157 are catalytic residues predicted for the second group. Remaining predicted amidase catalytic domains (endolysins from Raleigh, Darolandstone, Ibantik, Immanuel3) contain different amino acid residues or these sequences were predicted without strictly defined ends.

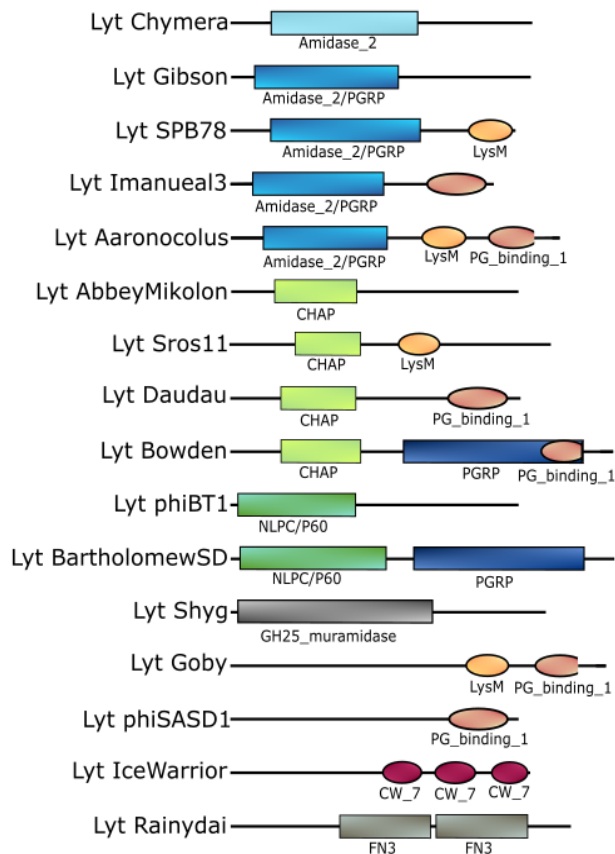
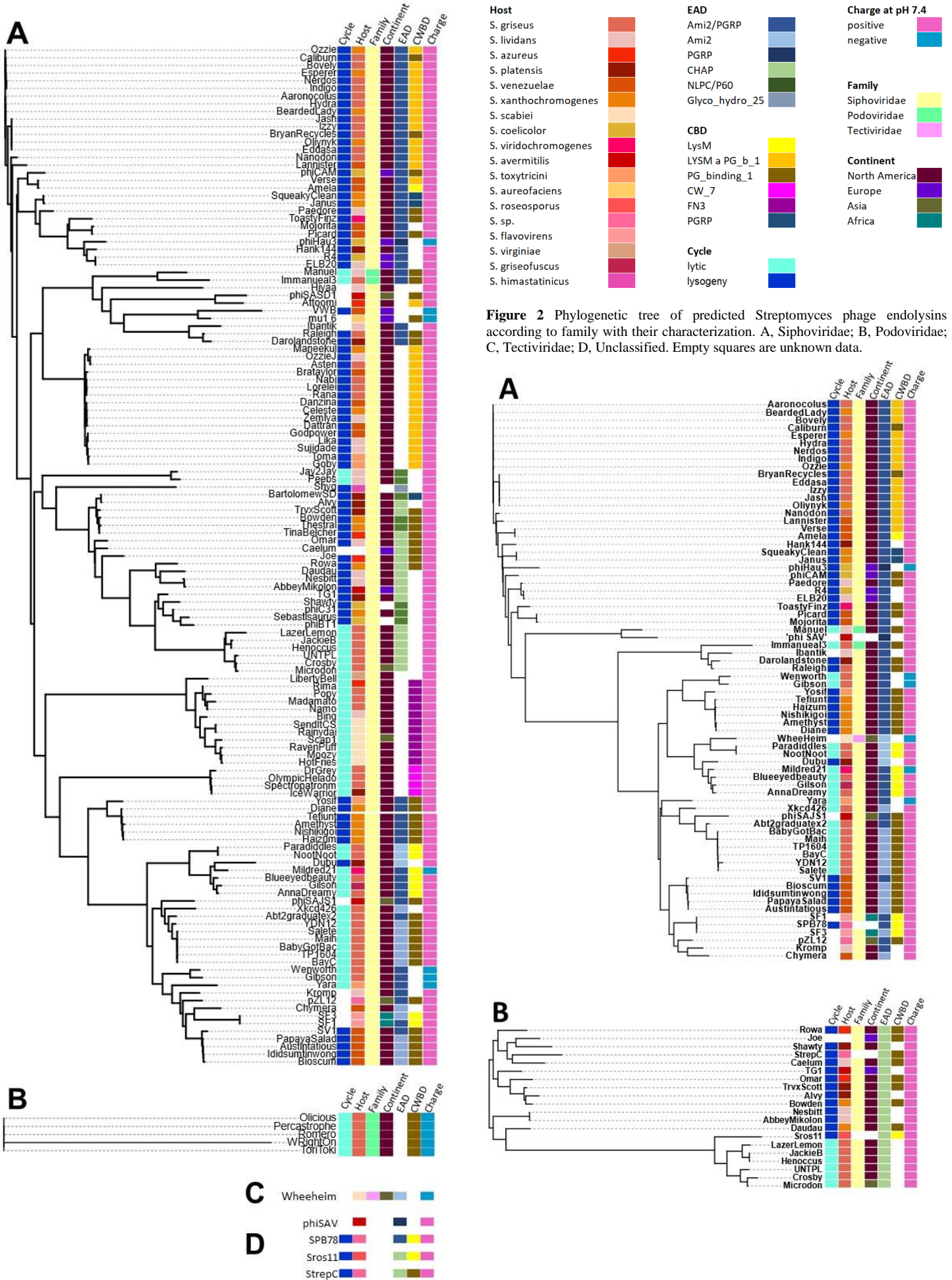


Figure 1 Schematic visualization of selected representative predicted domain structure



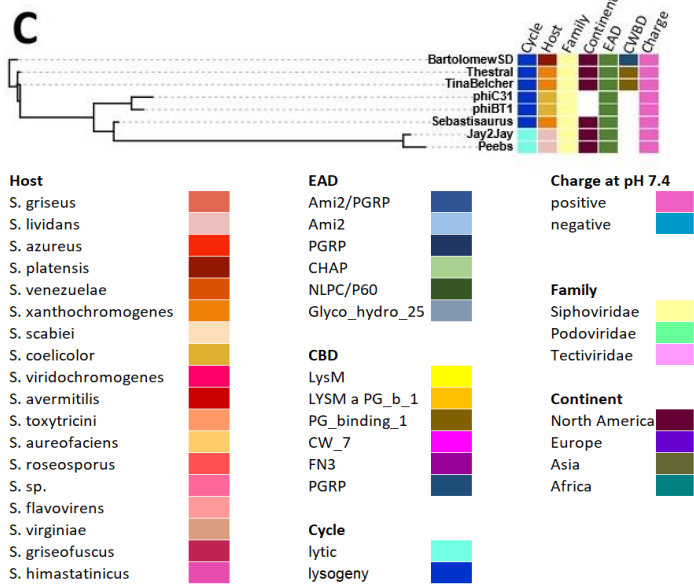


Figure 3 Phylogenetic tree of predicted *Streptomyces* phage endolysins their EAD with their characterization. A, Ami₂; B, CHAP; C, NLPC/P60. Empty squares are unknown data.

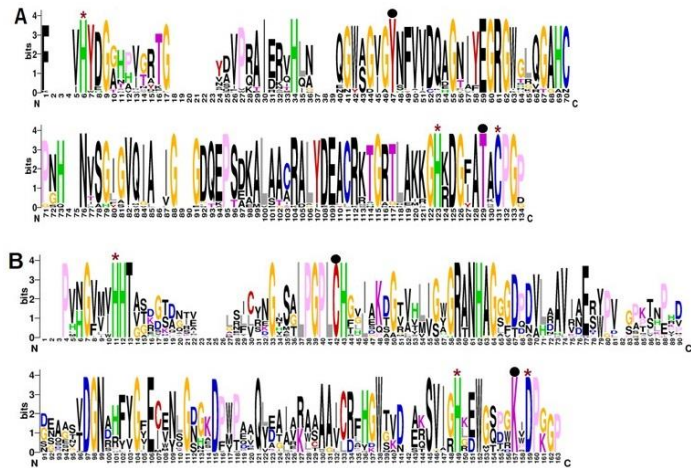


Figure 4 Weblogo of catalytic domains with predicted amidase activity (Amidase₂ and PGRP). Zinc binding residues are marked by red stars, catalytic residues by black dots. A: Domains with predicted Tyr47 and Thr129, as catalytic residues and zinc binding amino acid residues His6, His123, Cys131. B: Domains with predicted zinc binding residues His11, His149 and Asp159 and catalytic residues Cys42 and Lys157.

The Amidase₂ domain model (Figure 4) from endolysin phage Aaronocolus, a representative of the first group, Zn²⁺ binding residues (two histidines and cysteine) have the same position as in the template (PDB 4ZXM, *Branchiostoma belcheri tsingtauense* peptidoglycan recognition protein 3). Catalytic residues (tyrosine and tryptophan) are also in the same position but with different rotation.

The model therefore contributes to the accuracy of the prediction within the first group. In the model of Amidase₂ from endolysin phage SV1, a representative of the second group Zn²⁺ binding residues (two histidines and aspartic acid) match the template (4BOL.A, AmpDh2 from *Pseudomonas aeruginosa* complexed with pentapeptide) as well as catalytic the rest is lysine, but the second catalytic residue is not cysteine as in the weblogo. In the model, instead of cysteine, the second catalytic residue is glutamic acid, which is in the same position as Glu106 in the template.

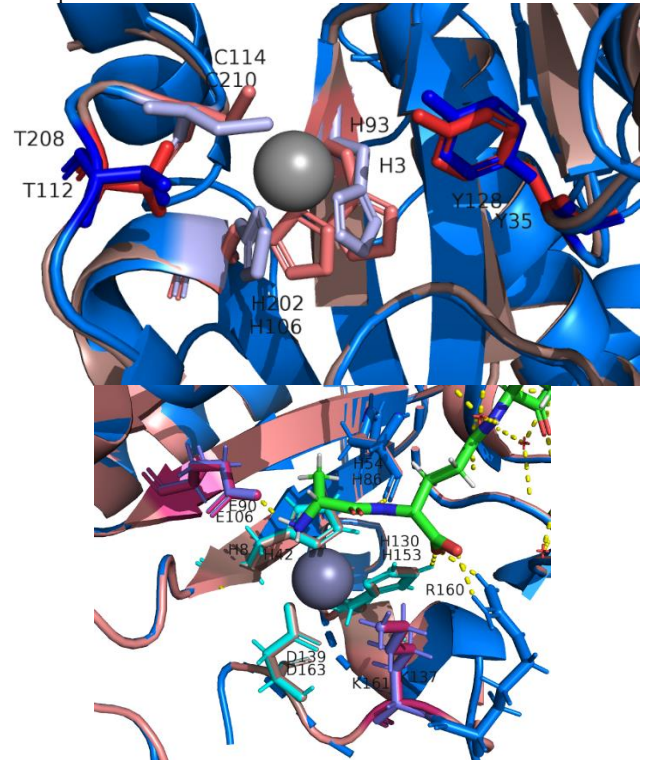


Figure 5 Left, model of the Amidase₂ domain of endolysin from Aaronocolus phage (red shades) with the crystal structure of *Branchiostoma belcheri tsingtauense* peptidoglycan recognition protein 3 (blue shades) and Zn²⁺ (PDB 4Z81). Right, model of the active site of the Amidase₂ domain of the SV1 phage endolysin (blue) with the AmpDh2 crystal structure from *Pseudomonas aeruginosa* complexed with pentapeptide and Zn²⁺ (salmon colour, PDB entry 4BOL.A).

Endolysins containing CHAP domain are only from phages of *Siphoviridae* family (Figure 3B) except endolysins from StrepC and Sros11 with unknown data. All of them have positive net charge and almost all contain PG_{binding_1} domain as cell wall binding domain. For this domain CD-search did not predict conserved catalytic amino acid residues, but alignment with CHAPk domain from LysK shows conserved Cys61 and His130 (Figure 6). The structure alignment of modelled Microdon endolysin did not show noticeable similarity with its template because of very low sequence identity (generally lower than 30%).

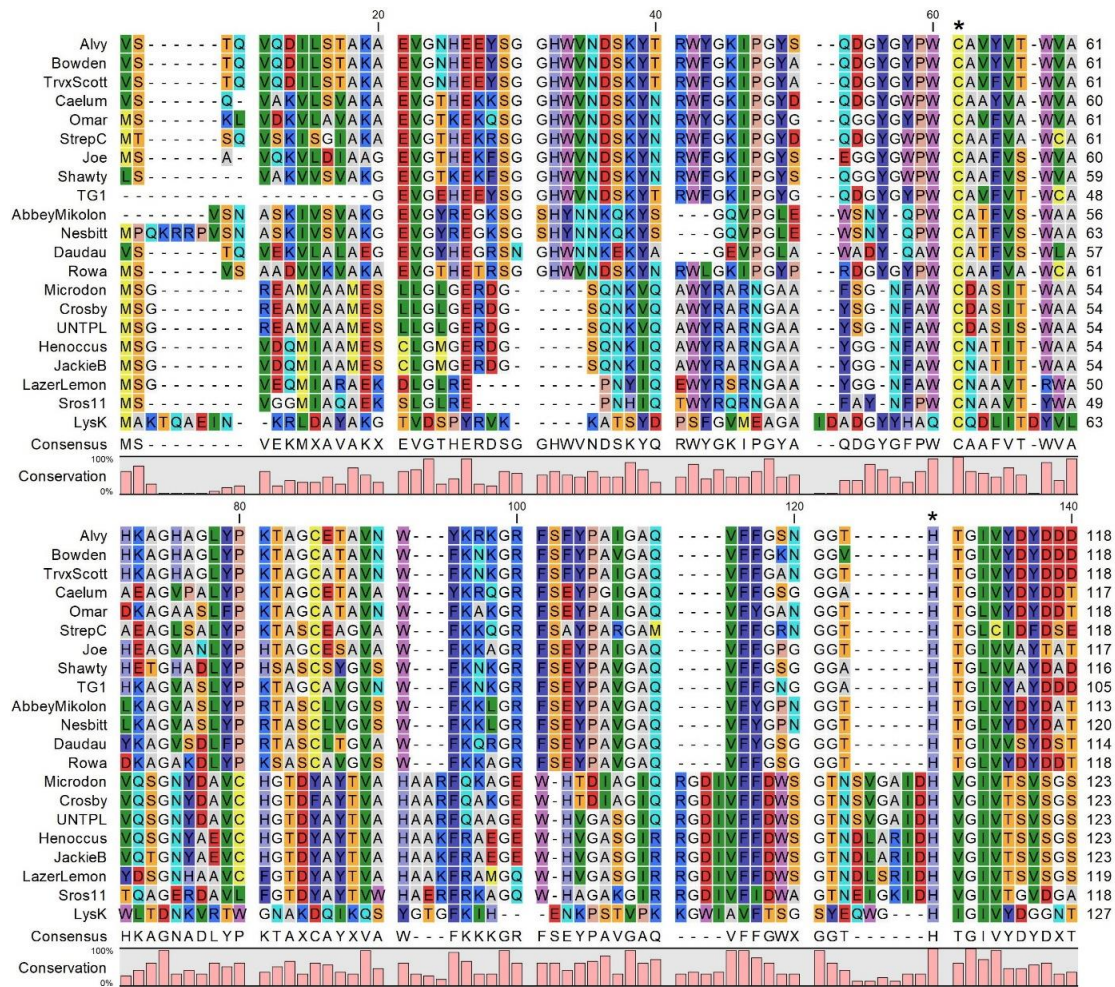


Figure 6 Part of alignment of N-terminus of predicted endolysins with CHAP domain and CHAPk domain of LysK. Conserved catalytic residues cysteine and histidine are marked with black stars

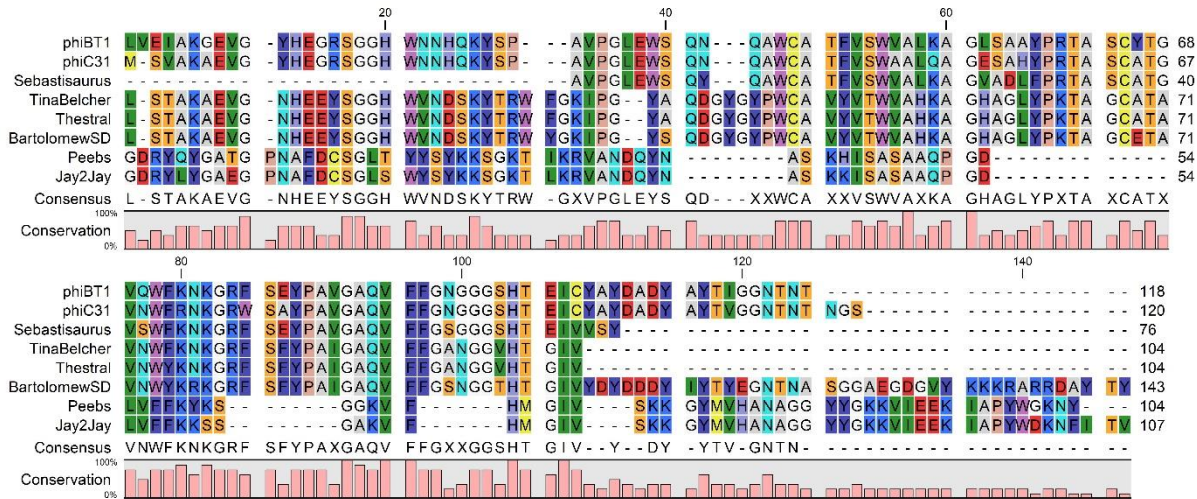


Figure 8 Alignment of predicted NlpC/P60 domain in predicted endolysins

Muramidases

Muramidases (EC 3.2.1.17) hydrolyze the 1,4-beta glycosidic bond between N-acetylmuramic acid and N-acetyl-D-glucosamine residues. Model of predicted Glyco_hydro_25 domain of Shyg phage endolysins has β/α -barrel fold like its template bacterial lysozyme from *Streptomyces coelicolor* (PDB entry 1JFX). We can see also conserved aspartic and glutamic acids (Figure 7).

Peptidases

The NlpC/P60 and CHAP domain nomenclature are sometimes used interchangeably because of their papain-like folding (Anantharaman and Aravind, 2003). The lytic activity of NlpC/P60 has been reported (Sheenan et al., 1997). Phages containing endolysins with predicted NlpC/P60 domains belongs only to *Siphoviridae* family and have positive net charge (Figure 3C). The conserved catalytic amino acids residues in predicted NlpC/P60 domains

were not predicted by CD-search neither by modelling nor by structure alignment. In alignment (Figure 8) we can see only good conserved sequences of these predicted domains and according to CwIT bifunctional hydrolase, in alignment are conserved residues Asp15 and Ser17 (positions in alignment) (Xu et al., 2014).

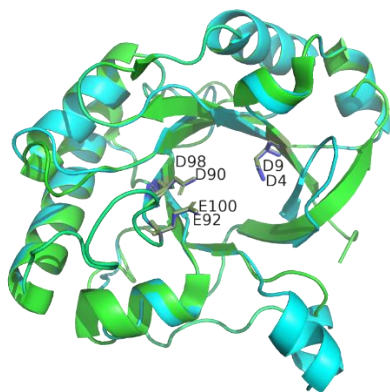


Figure 7 Structure alignment of Glyco_hydro_25 domain of Shyg phage endolysin (blue) with Crystal structure of the bacterial lysozyme from *Streptomyces coelicolor* (green), PDB entry 1JFX(8-193).

DISCUSSION

Phages and bacteria are in constant antagonist coevolution and *Streptomyces* phages undergo the developmental cycle with their hosts Streptomycetes because of their specific features like mycelium (Betts et al., 2014, Smith et al., 2013). Modular or non-modular structure of endolysins depends on host bacteria and for Gram-positive bacteria is typical modular structure with enzymatic active domain on N-terminus and cell wall binding domain on the C-terminus connected by a linker (Fenton et al., 2010; Schmelcher, Donovan, et al., 2012).

Set of 143 late endolysins was predicted in 175 *Streptomyces* phage genomes. The organization of genomes has typical mosaic structure as was described before for streptomycetes viruses (Smith et al., 2013), but in case of larger genomes, the mosaic structure is different. These genomes contain lot of insertions, non-typical proteins for this type of viruses and also predicted endolysins have not typical position after tail proteins and contain numerous types enzymes containing domains related to endolysins. So, it was harder to predict which enzyme is a typical late endolysin. Similar mosaic structure and anomalies were found in *Arthrobacter* viruses (Klyczek et al., 2017). For most of endolysins was predicted domain structure (Figure 1, Table 2) with N-terminal enzymatic active domain and C-terminal cell wall-binding domain, typical for virus endolysins infecting Gram-positive bacteria (Diaz et al., 1990). As enzymatic active domains were predicted Amidase_2 (PF01510), PGRP (cd06583), CHAP (PF05257), NlpC/P60 (PF00877) and in one case Glyco_hydro_25 (PF01183). On C-terminus as cell wall-binding domains were predicted LysM (PF01476), PG_binding_1 (PF01471), CW_7 (PF08230) and PGRP (cd06583). Moreover, fn3 domain (PF00041) was predicted near the holin, where endolysin is usually located, its function remains unclear (Valk et al., 2017), but it is predicted to have some role in glycoside hydrolases. Nearly all Gram-positive phage endolysins and autolysins are one gene product. Although group I introns are often found within these genes and have been reported for *Streptococcus* and *Staphylococcus* (Nelson et al., 2012). The net charge of membranes and bacterial cell walls is globally negative, logically most of antimicrobials and phage lytic enzymes display positive net charges (Low et al., 2011). Predicted endolysins in this study reflect on this observation, except few of them (10%). The negative net charge is specific for endolysins of phages from *Podoviridae* family. The reason for these exceptions can be the ability of some bacteria to modify their lipo-, exo- and capsular polysaccharide rides to sheer the human immune system (Valero, 2019). Life cycle of *Streptomyces* phages is balanced between only lytic (n=74, 42.3%) and lysogeny cycle (n=82, 46.6%). Geographically, most of studied phages origin from North America, which is caused by Phage Hunters Integrating Research and Education (PHIRE) program, where the phages are isolated and sequenced by novice high school and college scientists (Hanauer et al., 2006; Hatfull et al., 2006, Hatfull, 2015). Also, the establishment of the Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) program in 2008 (Jordan et al., 2014) increased isolation and sequencing of Actinobacteriophages.

Most of the endolysins have been predicted with amidase catalytic activity, where Ami_2 and PGRP are homologs defined in different databases. Prediction of Amidase_2 domain and PGRP domain by CD-search either varies in length (1-2 amino acids), position (shift by 2 amino acids) or they are identical. Predicted endolysins with FN3 CBD domain are also localized in this area, so it is possible that they contain modified Amidase_2 or PGRP domain. Amidase_2 have Zn²⁺ depending N-acetylmuramoyl-L-alanine amidase activity. In comparison, the two groups were clearly visible after marking the predicted amino acid residues located in the active site according to CD-search. The first contained as Zn²⁺ binding residues His6, His123 and Cys131 (positions in the weblogo) and catalytic residues Tyr47 and Thr19. In the second group, His11, His149, Asp159 and Cys42, Lys157 binding residues were predicted as Zn²⁺ binding residues. Exceptions to these two groups were the amidase domains of endolysins from the Raleigh, Darolandstone, Ibantik and Immanuel3 phages for which other

catalytic residues were predicted in the active site, or the domain boundaries were not clearly predicted for these sequences. The zinc-binding residues of the first group, two histidines and cysteine, coincide with already structurally confirmed amino acid residues, where zinc is coordinated by these residues in a tetrahedral formula (Gu et al., 2014). The second group, in which cysteine is replaced by aspartic acid, is the conformation that is possible (Zoll et al., 2010), but as mentioned above, has not yet been identified (Love et al., 2020). The second group would therefore require experimental verification. As for the catalytic residues, they only partially coincide with the above, glutamic acid and lysine, where it is believed that the water molecule as a nucleophile is activated by glutamic acid and lysine stabilizes the intermediate of the reaction, but instead of lysine there may be threonine at this position (Low et al., 2005). Thus, the alternation of lysine and threonine is identical to said endolysins. Such replacement has also been performed for bacteriophage T7 endolysin (Cheng et al., 1994), which contains the catalytic residues Tyr46 and Lys128, which could confirm the predicted residues within the first group.

When we look on models, the representant of the first group have in the same position as template zinc binding residues, two His and Cys and catalytic residues, Tyr and Trp. However, the representant of second group although have conserved zinc binding residues, two His and Asp are same as the template, but the second catalytic residue is not Cys. In the model, instead of Cys, the second catalytic residue is Glu, which is in the same position as Glu106 in the template. When we look back at the weblogo, we can see that glutamic acid is also fully preserved, so it is necessary to use experiments to verify which prediction is correct. In the past the only fully characterized endolysin was from actinophage Mu1/6 with amidase activity (Farkašová et al., 2016).

CHAP domain according to CD-search belongs to NLPC/P60 superfamily (Marchler-Bauer et al., 2017). In phylogenetic tree are located together and this cluster contains also endolysins with predicted CW_7 CBD. When we look at endolysin alignment part of phylogenetic tree from Peebs to phiBT1 we can see some similar patterns on N-terminus. The CHAP domain, cysteine histidine-dependent amidohydrolase / peptidase, contains about 120 amino acids with a unique potential to acquire two different enzymatic activities, amidohydrolase or peptidase (Bateman and Rawlings, 2003). As a peptidase, it cleaves the bonds between D-alanine and the first glycine in the pentaglycine bridge, and as an amidase, it cleaves the bond between the N-acetylmuramic acid residue and L-alanine at the N-terminus of the peptide (Bateman and Rawlings, 2003; Ridgen et al., 2003). Thus, proteins containing a CHAP domain can acquire only peptidase activity (e.g., LysK) or only amidase activity (e.g., Sk1), or both enzymatic activities (e.g., LytN) (Vermassen et al., 2019).

Cysteine, Histidine-dependent Amidohydrolases/Peptidase (CHAP) superfamily is involved in cell wall hydrolysis (Xu et al., 2009). This superfamily contains two families according to CD-search, the CHAP family (PF05257) and the NLPC/P60 family (PF00877). In InterPro they belong also in Papain-like cysteine peptidase superfamily IPR038765. Both domains contain strictly conserved cysteine and histidine residues and have the length about 110-140 amino acid residues. The NlpC/P60 or CHAP domains are widespread in bacteria and members of this superfamily have also been detected in bacteriophages, viruses, archaea and eukaryotes (Anantharaman and Aravind, 2003). NlpC/P60 domain forms multifunctional proteins with other components, such as LysM, SH3 and choline-binding domains (Smith et al., 2000). In Pfam database we can find almost 300 different architecture of this domain.

Glyco_hydro_25, glycoside hydrolase family 25, glycoside hydrolases EC 3.2.1 are a widespread group of enzymes that hydrolyze the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety (<https://pfam.xfam.org/family/PF01183>). In CAZY database this family comprises enzymes with only lysozyme activity (https://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_25). In this study was Glyco_hydro predicted only for one endolysin encoded by Shyg phage and according to modelling takes on similar conformation as lysozyme from *Streptomyces coelicolor*.

The results of this study show the diversity of Streptmyces phages and their endolysins relations with host specificity, and origin. The *in silico* characterization of domain, primary and tertiary structure spreads knowledge about these endolysins and this information can be used for developing new types of endolysins, enzybiotics by domain shuffling, deletion, target site mutations. Addition of whole domains or parts, e. g. amphipathic regions from endolysins coded by different phages can result in new chimeric endolysins with novel properties that can be either extended or species-specific, strain-specific, or serovar-specific.

Acknowledgment: This work is supported by grant APVV-16-0173.

REFERENCES

Anantharaman V, Aravind L. (2003). Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome biology*, 4(2), R11. <http://dx.doi.org/10.1186/gb-2003-4-2-r11>

- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*, 44(W1), W16-21. <http://dx.doi.org/10.1093/nar/gkw387>
- Bai, J., Kim, Y.-T., Ryu, S., and Lee, J.-H. (2016). Biocontrol and Rapid Detection of Food-Borne Pathogens Using Bacteriophages and Endolysins. *Frontiers in microbiology*, 7, 474. <http://dx.doi.org/10.3389/fmicb.2016.00474>
- Bateman, A., Bycroft, M. (2000). The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *Journal of molecular biology*, 299(4), 1113-9. <http://dx.doi.org/10.1006/jmbi.2000.3778>
- Bateman, A., Rawlings, N. D. (2003). The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends Biochem. Sci.* 28, 234–237. [http://dx.doi.org/10.1016/S0968-0004\(03\)00061-6](http://dx.doi.org/10.1016/S0968-0004(03)00061-6)
- Betts, A., Kaltz, O., Hochberg, M.E. (2014). Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages. *Proceedings of the National Academy of Sciences of the United States of America*, 111(30), 11109-14. <http://dx.doi.org/10.1073/pnas.1406763111>
- Cheng, Q., Fischetti, V. A. (2007). Mutagenesis of a bacteriophage lytic enzyme PlyGBS significantly increases its antibacterial activity against group B streptococci. *Applied Microbiology and Biotechnology*, 74(6), 1284-91. <http://dx.doi.org/10.1007/s00253-006-0771-1>
- Cheng, X., Zhang, X., Pflugrath, J. W., Studier, F. W., (1994). The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, 91(9), 4034-8. <http://dx.doi.org/10.1073/pnas.91.9.4034>
- Crooks, G. E., Hon, G., Chandonia, J. M., Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome research*, 14(6), 1188-90. <http://dx.doi.org/10.1101/gr.849004>
- Farkašová, J., Godány, A. (2016). Characterization of the N-Terminal Catalytic Domain of Lyl1/6, an Endolysin from *Streptomyces aureofaciens* Phage μ 1/6. *Current microbiology*, 73(4), 602-10. <http://dx.doi.org/10.1007/s00284-016-1100-2>
- Fenton, M., Ross, P., McAuliffe, O., O'Mahony, J., Coffey, A. (2010). Recombinant bacteriophage lysins as antibacterials. *Bioengineered bugs*, 1(1): 9–16. <http://dx.doi.org/10.4161/bbug.1.1.9818>
- Fischetti, V. A. (2008). Bacteriophage lysins as effective antibacterials. *Current opinion in microbiology*, 11(5), 393-400. <http://dx.doi.org/10.1016/j.mib.2008.09.012>
- García, J. L., García, E., Arrarás, A., García, P., Ronda, C., López, R. (1987). Cloning, purification, and biochemical characterization of the pneumococcal bacteriophage Cp-1 lysin. *Journal of virology*, 61(8):2573-80.
- Gómez-Torres, N., Dunne, M., Garde, S., Meijers, R., Narbad, A., Ávila, M., and Mayer, M. J. (2018). Development of a specific fluorescent phage endolysin for in situ detection of *Clostridium* species associated with cheese spoilage. *Microbial biotechnology*, 11(2), 332-345. <http://dx.doi.org/10.1111/1751-7915.12883>
- Gu, J., Feng, Y., Feng, X., Sun, C., Lei, L., Ding, W., Niu, F., Jiao, L., Yang, M., Li, Y., Liu, X., Song, J., Cui, Z., Han, D., Du, C., Yang, Y., Ouyang, S., Liu, Z., Han, W. (2014). Structural and biochemical characterization reveals LysGH15 as an unprecedented "EF-hand-like" calcium-binding phage lysin. *PLoS Pathog.* 10 (2014) e1004109. <http://dx.doi.org/10.1371/journal.ppat.1004109>
- Hanauer, D. I., Jacobs-Sera, D., Pedulla, M. L., Cresawn, S. G., Hendrix, R. W., Hatfull G. F. (2006). Inquiry learning. Teaching scientific inquiry. *Science*, 314(5807), 1880-1. <http://dx.doi.org/10.1126/science.1136796>
- Hatfull, G. F. (2015). Innovations in Undergraduate Science Education: Going Viral. *Journal of virology*, 89(16), 8111-3. <http://dx.doi.org/10.1128/JVI.03003-14>
- Hatfull, G. F., Pedulla, M. L., Jacobs-Sera, D., Cichon, P. M., Foley, A., Ford, M. E., Gonda, R. M., Houtz, J. M., Hryckowian, A. J., Kelchner, V. A., Namburi, S., Pajcini, K. V., Popovich, M. G., Schleicher, D. T., Simanek, B. Z., Smith, A. L., Zdanowicz, G. M., Kumar, V., Peebles, C. L., Jacobs, W. R. Jr, Lawrence, J. G., Hendrix, R. W. (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS genetics*, 2(6), e92. <http://dx.doi.org/10.1371/journal.pgen.0020092>
- Hoopes, J.T., Stark, C.J., Kim, H.A., Sussman, D.J., Donovan, D.M., and Nelson, D.C. (2009) Use of a bacteriophage lysin, PlyC, as an enzyme disinfectant against *Streptococcus equi*. *Applied and Environmental Microbiology*, 75(5), 1388-94. <http://dx.doi.org/10.1128/AEM.02195-08>
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends in biochemical sciences*, 23(10), 403-5. [http://dx.doi.org/10.1016/S0968-0004\(98\)01285-7](http://dx.doi.org/10.1016/S0968-0004(98)01285-7)
- Jones, D. T., Taylor, W. R., Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS*, 8(3):275-82. <http://dx.doi.org/10.1093/bioinformatics/8.3.275>
- Jordan, T. C., Burnett, S. H., Carson, S., Caruso, S. M., Clase, K., DeJong, R. J., Dennehy, J. J., Denver, D. R., Dunbar, D., Elgin, S. C., Findley, A. M. et al. (2014). A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *mBio*, 5(1), e01051-13. <http://dx.doi.org/10.1128/mBio.01051-13>
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Stenberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6), 845-58. <http://dx.doi.org/10.1038/nprot.2015.053>
- Kerff, F., Petrella, S., Mercier, F., Sauvage, E., Herman, R., Pennartz, A., Zervosen, A., Luxen, A., Frère, J. M., Joris, B., Charlier, P. (2010). Specific structural features of the N-acetylmuramoyl-L-alanine amidase AmiD from *Escherichia coli* and mechanistic implications for enzymes of this family. *Journal of molecular biology*, 397(1), 249-59. <http://dx.doi.org/10.1016/j.jmb.2009.12.038>
- Klyczek, K. K., Bonilla, J. A., Jacobs-Sera, D., Adair, T. L., Afram, P., Allen, K. G., et al. (2017) Tales of diversity: Genomic and morphological characteristics of forty-six *Arthrobacter* phages. *PLoS One*, 12(7), e0180517. <http://dx.doi.org/10.1371/journal.pone.0180517>
- Kretzer, J. W., Lehmann, R., Schmelcher, M., Banz, M., Kim, K.-P., Korn, C., and 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.org/10.1128/AEM.02402-06>
- Letunic, I., Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research*, 47(W1), W256-W259. <http://dx.doi.org/10.1093/nar/gkz239>
- Loeffler, J. M., Nelson, D., Fischetti, V. A. (2001). Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. *Science*, 294(5549), 2170-2. <http://dx.doi.org/10.1126/science.1066869>
- Love, M. J., Abeysekera, G. S., Muscroft-Taylor, A. C., Billington, C., Dobson, R. C. J. (2020). On the Catalytic Mechanism of Bacteriophage Endolysins: Opportunities for Engineering. *Biochim Biophys Acta Proteins Proteom*, 1868(1), 140302. <http://dx.doi.org/10.1016/j.bbapap.2019.140302>
- Low, L.Y., Yang, C., Perego, M., Osterman, A., and Liddington, R.C. (2005). Structure and lytic activity of a *Bacillus anthracis* prophage endolysin. *The journal of biological chemistry*, 280(42):35433-9. <http://dx.doi.org/10.1074/jbc.M502723200>
- Low, L.Y., Yang, C., Perego, M., Osterman, A., Liddington, R. (2011). Role of net charge on catalytic domain and influence of cell wall binding domain on bactericidal activity, specificity, and host range of phage lysins. *The journal of biological chemistry*, 286(39), 34391-403. <http://dx.doi.org/10.1074/jbc.M111.244160>
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic acids research*, 43, D222-6. <http://dx.doi.org/10.1093/nar/gku1221>
- Nelson, D., Loomis, L., Fischetti, V. A. (2001). Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 98(7), 4107-12. <http://dx.doi.org/10.1073/pnas.061038398>
- Nelson, D. C., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D. G., Dong, S., Donovan, D. M. (2012). Endolysins as antimicrobials. *Advances in virus research*, 83, 299-365. <http://dx.doi.org/10.1016/B978-0-12-394438-2.00007-4>
- Nelson, D., Schuch, R., Chahales, P., Zhu, S., Fischetti, V. A. (2006). PlyC: A multimeric bacteriophage lysin. *Proceeding of the Academy of Sciences of the United States of America*, 103(28), 10765-70. <http://dx.doi.org/10.1073/pnas.0604521103>
- Nelson, D., Schuch, R., Zhu, S., Tscherne, D. M., Fischetti, V. A. (2003). Genomic sequence of C1, the first streptococcal phage. *Journal of Bacteriology*, 185(11), 3325-32. <http://dx.doi.org/10.1128/jb.185.11.3325-3332.2003>
- Oliveira, H.; Azeredo, J.; Lavigne, R.; Kluskens, L.D. (2012). Bacteriophage endolysins as a response to emerging foodborne pathogens. *Trends in food science & technology*, 28(2), 103-15. <http://dx.doi.org/10.1016/j.tifs.2012.06.016>
- Rigden, D. J., Jedrzejak, M. J., and Galperin, M. Y. (2003). Amidase domains from bacterial and phage autolysins define a family of gamma-D,L-glutamate-specific amidohydrolases. *Trends Biochem. Sci.* 28, 230–234. [http://dx.doi.org/10.1016/S0968-0004\(03\)00062-8](http://dx.doi.org/10.1016/S0968-0004(03)00062-8)
- Saitou, N., Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4), 406-25. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a040454>
- Schmelcher, M., Donovan, D. M., Loessner, M. J. (2012). Bacteriophage endolysins as novel antimicrobials. *Future microbiology*, 7(10), 1147-71. <http://dx.doi.org/10.2217/fmb.12.97>
- Schmelcher, M., and Loessner, M.J. (2016). Bacteriophage endolysins: applications for food safety. *Current opinion in biotechnology*, 37, 76-87. <http://dx.doi.org/10.1016/j.copbio.2015.10.005>
- Schmelcher, M., Powell, A.M., Camp, M.J., Pohl, C.S., Donovan, D. M. (2015). Synergistic streptococcal phage λ SA2 and B30 endolysins kill streptococci in cow milk and in a mouse model of mastitis. *Applied microbiology and biotechnology*, 99(20), 8475–8486. <http://dx.doi.org/10.1007/s00253-015-6579-0>
- Sheenan, N. M., García, J. L., López, R., García, P. (1997). The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin. *Molecular*

- microbiology, 25(4), 717-25. <http://dx.doi.org/10.1046/j.1365-2958.1997.5101880.x>
- Smith, T. J., Blackman, S. A., Foster, S. J. (2000). Autolysins of *Bacillus subtilis*: multiple enzymes with multiple functions. *Microbiology*, 146(Pt 2), 249-262. <http://dx.doi.org/10.1099/00221287-146-2-249>
- Smith, M. C., Hendrix, R. W., Dedrick, R., Mitchell, K., Ko, C. C., Russell, D., Bell, E., Gregory, M., Bibb, M. J., Pethick, F., Jacobs-Sera, D., Herron, P., Buttner, M. J., Hatfull, G. F. (2013). Evolutionary relationships among actinophages and a putative adaptation for growth in *Streptomyces* spp. *Journal of bacteriology*, 195(21), 4924-35. <https://dx.doi.org/10.1128/JB.00618-13>
- Valero-Rello, A. (2019). Diversity, specificity and molecular evolution of the lytic arsenal of *Pseudomonas* phages: in silico perspective. *Environmental Microbiology*, 21(11), 4136-4150. <https://dx.doi.org/10.1111/1462-2920.14767>
- Valk V., Van der Kaaij M., R., Dijkhuizen, L. (2017). The evolutionary origin and possible functional role of FNIII domains in two microbacterium aurum B8.A granular starch degrading enzymes, and in other carbohydrate acting enzymes. *Amylase*, 1(1), 1-11. <https://dx.doi.org/10.1515/amylase-2017-0001>
- Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R., Wishart, D. S. (2005). BASys: a web server for automated bacterial genome annotation. *Nucleic acids research*, 33, W455-9. <https://dx.doi.org/10.1093/nar/gki593>
- Vermassen, A., Leroy, S., Talon, R., Provot, C., Popowska, M., Desvaux, M., (2019). Cell Wall Hydrolases in Bacteria: Insight on the Diversity of Cell Wall Amidases, Glycosidases and Peptidases Toward Peptidoglycan. *Front Microbiol.* 28. <https://dx.doi.org/10.3389/fmicb.2019.00331>
- Voelker, R. (2019). FDA Approves Bacteriophage Trial. *JAMA*, 321(7), 638. <https://dx.doi.org/10.1001/jama.2019.0510>
- Xu, Q., Sudek, S., McMullan, D., Miller, M. D., Geierstanger, B., Jones, D. H., Krishna, S. S., Spraggon, G., et al. (2009). Structural basis of murein peptide specificity of a gamma-D-glutamyl-L-diamino acid endopeptidase. *Structure*, 17(2), 303-13. <https://dx.doi.org/10.1016/j.str.2008.12.008>
- Young, R. (1992). Bacteriophage lysis: mechanism and regulation. *Microbiological reviews*, 56(3), 430-81.
- Zoll, S., Pätzold, B., Schlag, M., Götz, F., Kalbacher, H., Stehle, T. (2010). Structural basis of cell wall cleavage by a staphylococcal autolysin. *PLoS Pathog*, 6(3), e1000807. <https://dx.doi.org/10.1371/journal.ppat.1000807>
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic acids research*, 39(2), W347-52. <https://dx.doi.org/10.1093/nar/gkr485/>