A Novel Strategy to Identify Endolysin With Lytic Activity Against Methicillin-Resistant *Staphylococcus aureus*

Hanbeen Kim [1] and Jakyeom Seo [1,*]

[1] Department of Animal Science, Life and Industry Convergence Research Institute, Pusan National University, Miryang 50463 Republic of Korea; khb3850@pusan.ac.kr and jseo81@pusan.ac.kr

* Correspondence: jseo81@pusan.ac.kr; Tel.: +82-55-350-5513; Address: Department of Animal Science, College of Natural Resources & Life Sciences, Pusan National University, 1268-50 Samrangjin-ro, Miryang 50463, Republic of Korea.

Abstract: The increasing prevalence of methicillin-resistant *S. aureus* (MRSA) in the dairy industry has become a fundamental concern. Endolysins are bacteriophage-derived peptidoglycan hydrolases that induce the rapid lysis of host bacteria. We investigated endolysin candidates with lytic activity against MRSA and evaluated the lytic activity of the endolysin candidate against *S. aureus* and MRSA. To identify endolysins, we used the following bioinformatic strategy: (1) retrieval of genetic information, (2) annotation, (3) selection of MRSA, (4) selection of endolysin candidates, and (5) evaluation of protein solubility. We then characterized the endolysin candidate under various conditions. Approximately 67% of *S. aureus* was detected as MRSA and a total of 114 putative endolysins were found. The 114 putative endolysins were divided into three groups based on their combination of conserved domains. Considering the protein solubility, we selected putative endolysins 177 and 117. Putative endolysin 117 was successfully overexpressed and renamed LyJH1892. LyJH1892 lysed approximately 85% of *S. aureus*. LyJH1892 showed more potent lytic activity against MRSA than normal *S. aureus* (relative lytic activity > 125%). LyJH1892 showed broad lytic activity against coagulase-negative Staphylococci (CNS). In conclusion, These findings provide a rapid and useful strategy for the development of specific endolysins against antibiotic-resistant bacterial strains.

Keywords: Prophage; Endolysin; *Staphylococcus aureus*; Methicillin-resistant *Staphylococcus aureus*; Antibiotic-resistance; Bovine mastitis.

1. Introduction

Staphylococcus aureus is a gram-positive bacterium and is a part of the normal microbiota residing on the skin and nares of animals and humans [1,2]. *S. aureus* is considered an opportunistic pathogen involved in a variety of infectious diseases, such as skin infections, food poisoning, toxic shock syndrome, and endocarditis [3,4]. In addition, the treatment of *S. aureus* infections has been challenging because of the emergence of antibiotic-resistant *S. aureus*, such as methicillin-resistant *S. aureus* (MRSA), vancomycin-resistant *S. aureus*, and multidrug-resistant *S. aureus* [5,6]. In the dairy industry, *S. aureus* is considered one of the most concerning, contagious pathogens that can cause bovine mastitis. This disease triggers an inflammatory response in the udder tissue, causing enormous economic losses in dairy farms [7]. A recent study reported that MRSA is a critical risk factor for bovine mastitis, representing 12.2% of all mastitis cases [8]. In addition, evidence has emerged suggesting direct transmission of MRSA from cows to humans working in the dairy industry [9]. Considering the poor efficacy of β-lactam antibiotics against MRSA, alternative strategies are required to combat this pathogen in dairy farms.

Endolysin is a bacteriophage-derived peptidoglycan hydrolase that induces the lysis of host bacteria by degrading the peptidoglycan layer. Therefore, this is a promising candidate for replacing antibiotics [10,11]. Endolysins active against gram-positive bacteria have generally shown multiple combinations of conserved domains, especially the N-

 doi:10.20944/preprints202302.0072.v1

terminal enzymatically active domain (EAD) and the C-terminal cell wall-binding domain (CBD) [11]. The use of endolysins as alternatives to antibiotics has been highlighted because they elicit low resistance rates, specific lysis, and rapid reaction [12].

There have been many efforts to develop anti-staphylococcal agents such as bacteriocin [13-15] and endolysin [16-18], and a recent review highlighted their potential as therapeutic agents against antibiotic-resistant *S. aureus* [1]. A general strategy to develop endolysins starts from the isolation of bacteriophages with lytic activity against host bacteria; however, the isolation of specific bacteriophages with lytic activity against a pathogen of interest is laborious and time-consuming. Recent studies have suggested a new strategy to develop a recombinant endolysin using the genomic information of prophages [19-21]. In this study, we established a strategy to identify and select endolysin candidates from genome information using several steps: (1) retrieval of *S. aureus* genome sequences, (2) annotation, (3) selection of MRSA containing penicillin-binding protein (PBP) 2a, (4) selection of endolysin genes candidates, and (5) evaluation of the predicted protein solubility.

This study describes the development of a recombinant endolysin using the MRSA prophage genome. Furthermore, we hypothesized that this endolysin has lytic activity against *S. aureus* and MRSA. Thus, the objective of this study was to investigate endolysin candidates from MRSA prophages and develop new endolysins with lytic activity against *S. aureus* and MRSA.

2. Results

2.1. Identification of putative endolysins from MRSA

A total of 707 complete genomes of *S. aureus* were retrieved from the NCBI database (retrieval date: 06.02.2020), and 474 of the complete *S. aureus* genomes exhibited PBP2a (Figure 1a and Table S1). In the MRSA prophage sources, 1,907 sequences were detected as endolysins; however, only 272 sequences remained for further analysis after removing sequences containing duplicate and/or incomplete conserved domains (Figure 1b and Table S2). Among 272 sequences of putative endolysins, 42.3% had a single domain (n = 115) and the others exhibited multi-conserved domain combinations (n = 157) (Figure 1c). The most frequent domain among the putative endolysins with a single domain was the LytD domain (approximately 52.2%, Accession: COG4193), followed by the cysteine, histidine-dependent amidohydrolase/peptidase (CHAP) domain (approximately 19.1%, Accession: pfam05257), and the SH3b domain (approximately 13.9%, Accession: smart00287) (Figure 1d). The most frequently observed multi-conserved domains were [CHAP (Accession: pfam05257) + Amidase_2 (Accession: smart00644) + SH3_5 (Accession: pfam08460)], followed by [CHAP (Accession: pfam05257) + Amidase_3 (Accession: pfam01520) + SH3b (Accession: smart00287)] and [CHAP (Accession: pfam05257) + LytD (Accession: COG4193)] (Figure 1e). Interestingly, among 157 putative endolysin sequences with multi-conserved domain combinations, only 114 putative endolysin sequences showed both EAD and CBD, indicating general characteristics for endolysins targeting gram-positive bacteria (Table S2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0072.v1

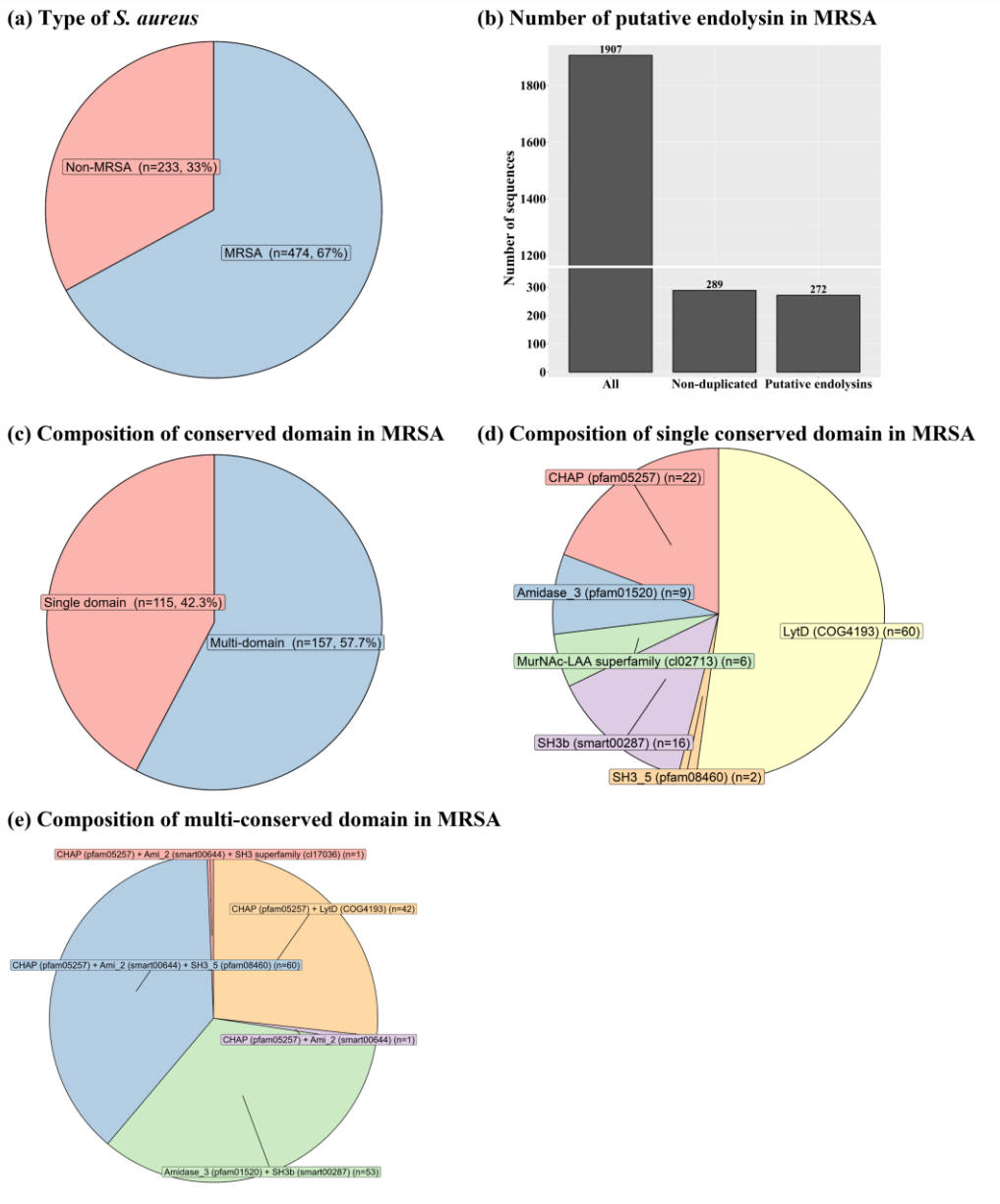

Figure 1. General information of *Staphylococcus aureus* genomes and putative endolysins. (a) Distribution of *S. aureus* and methicillin-resistant *S. aureus* (MRSA). (b) Numbers of putative endolysins observed in the MRSA genomes. (c) Distribution of domain types (single versus multiple domains). (d) Type of domains observed in single domain putative endolysins. (e) Type of domain combinations observed in multiple domain putative endolysins. The conserved domain was analyzed based on the conserved domain database from the National Center for Biotechnology Information.

2.2. Sequence comparison of putative endolysins targeting MRSA

The 114 putative endolysins with both EAD and CBD were divided into three groups based on their shared conserved domain combinations (groups 1, 2, and 3). The overall distributions of the 114 putative endolysins, based on the multi-alignment results and phylogenetic distances, were visualized using a cladogram plot (Figure 2a). Among the three putative endolysin groups, putative endolysins belonging to groups 1 and 2 showed not only close phylogenetic relationships belonging to the same clade but also high sequence identity (78.2%, data not shown), sharing CHAP (pfam05257) and Amidase_2 (smart00644) domains (Figure 2b). However, putative endolysins from group 3 belonged to different clades compared to groups 1 and 2 (Figure 2a), sharing low sequence identity (20.6%, data not shown). Interestingly, putative endolysins belonging to group 3 were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0072.v1

more conserved within the group (87.8%) than those belonging to group 1 (80.5%) (data not shown).

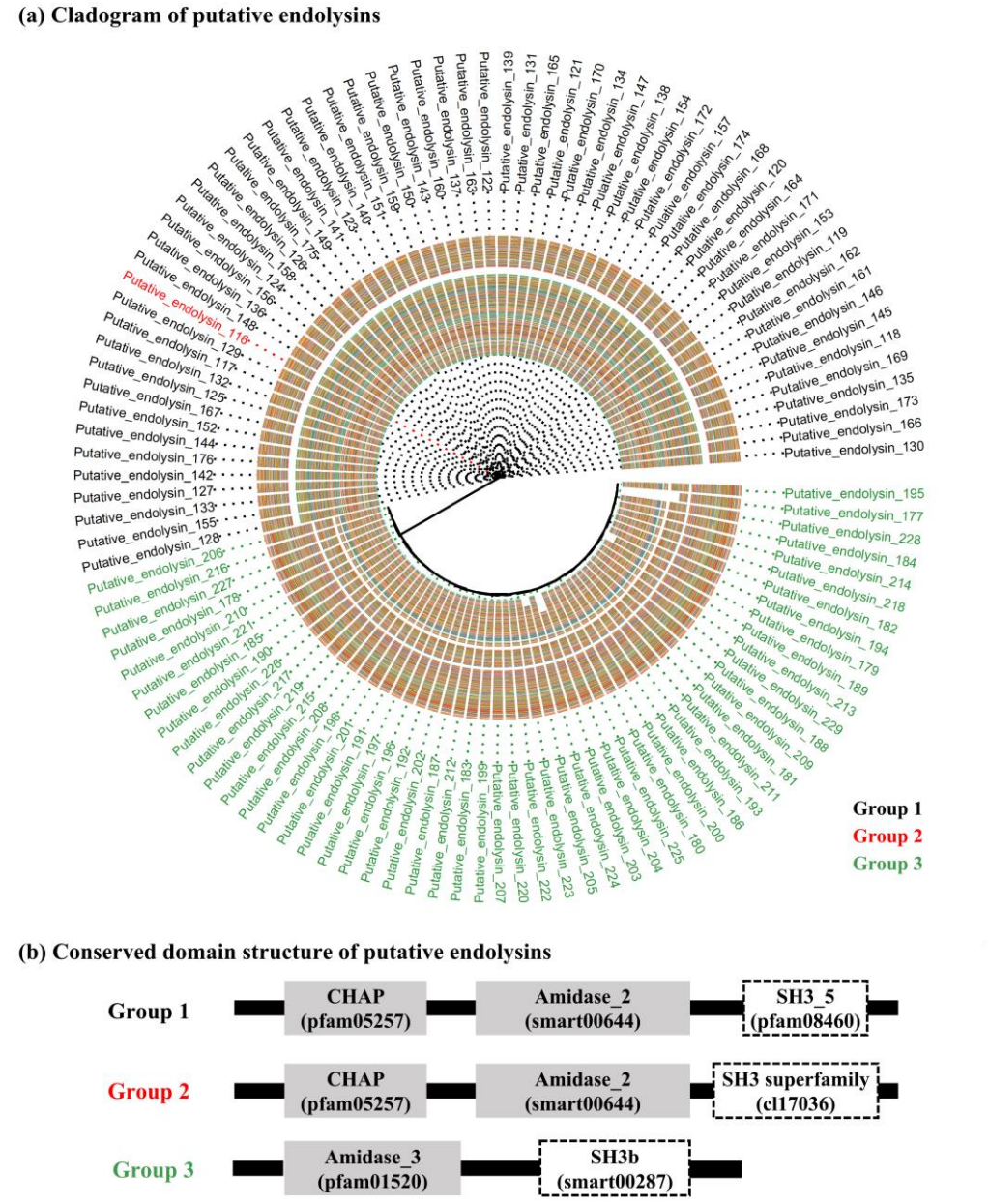

Figure 2. Novel diversity of putative endolysins having both enzymatically active domains and cell wall binding domains identified from methicillin-resistant *Staphylococcus aureus*-related genomes. (a) Cladogram of putative endolysins. A total of 114 putative endolysins were multi-aligned based on the amino acid sequences using the E-INS-I algorithm on multiple alignment using fast Fourier transform (version 7.505). The resulting alignments were phylogenetically reconstructed using FastTree, and the cladogram was constructed using ggtree in R software (version 4.1.3). Color tiles in the cladogram represent multi-alignment results among 114 putative endolysins. (b) Conserved domain structure of putative endolysins. CHAP, cysteine, histidine-dependent amidohydrolases/peptidases.

According to the multi-alignment of each conserved domain from the putative endolysins in group 1, CHAP and Amidase_2 domains were highly conserved (CHAP domain, sequence identity: 83.5% and Amidase_2 domain, sequence identity: 82.4%), whereas the SH3_5 domain exhibited slightly lower sequence similarity (sequence identity: 75.8%)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0072.v1

(Figure S1). In group 2, both Amidase_3 and SH3_b domains exhibited higher sequence identity than any of the sequence identities in group 1 (Amidase_3 domain, sequence identity: 95.1% and SH3_b domain, sequence identity: 88.4%) (Figure S2). Figure 3 shows overall comparisons of each conserved domain sharing a similar name and/or function from the 114 putative endolysins with EAD and CBD. Interestingly, the Amidase-like domain (Amidase_2 and Amidase_3) and SH3-like domain (SH3_5, SH3 superfamily, and SH3b) showed extremely low sequence identity (amidase-like domain, 12.7% and SH3-like domain, 22.9%), whereas the CHAP domain exhibited high sequence identity (82.4%) (Figure 3).

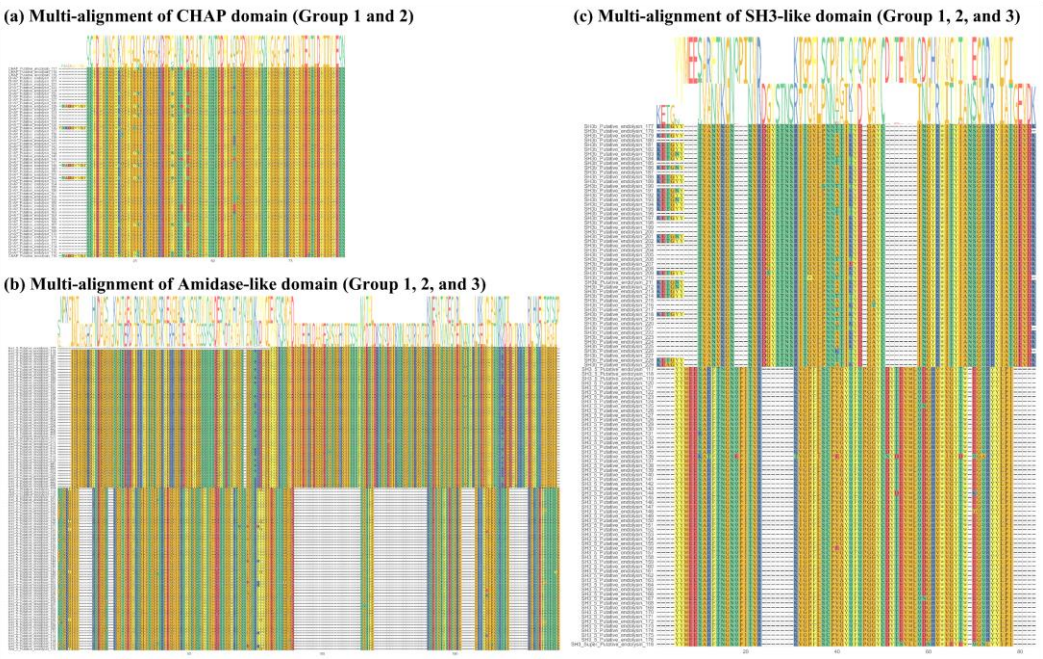

Figure 3. Multi-alignment of all putative endolysins with an enzymatically active domain and cell wall binding domain based on amino acid sequences. (a) Multi-alignment of the cysteine, histidine-dependent amidohydrolases/peptidases domain from groups 1 and 2 (Accession: pfam05257 and sequence identity: 82.4%). (b) Multi-alignment of Amidase-like domain (Accession: Ami_2 [smart00644] and Amidase_3 [pfam01520]; sequence identity: 12.7%). (c) Multi-alignment of the SH3-like domain (Accession: SH3_5 [pfam08460], SH3 superfamily [cl17036], and SH3b [smart00287]; sequence identity: 22.9%).

2.3. Comparison of predicted recombinant protein solubility

The predicted recombinant protein solubility based on three different bioinformatics tools (Soluprot, Protein-Sol, and SKADE) is presented in Table 1 and Table S3. Because of the different ranges of predicted solubility scores depending on the prediction tools, we first normalized each of the predicted solubility scores, that is, the solubility from Soluprot and SKADE, and scaled-sol from Protein-Sol. Additionally, the normalized scores were summed up to compare the predicted solubilities. Among the top 10 putative endolysins, 90% belonged to group 1, whereas the highest one was putative endolysin 177, which belongs to group 3. Based on the predicted solubility, we selected two putative endolysins, putative endolysin 177 (group 3) and putative endolysin 117 (group 3). Furthermore, we developed recombinant proteins to evaluate lytic activity against *S. aureus*.

Table 1. Top 10 of predicted recombinant protein solubility of putative endolysins based on Soluprot, Protein-Sol, and SKADE[1]

 doi:10.20944/preprints202302.0072.v1

Group[2]	Name	Amino acids	Soluprot		Protein-Sol		SKADE		Sum[7]
			Solubility[3]	Z-score[4]	Scaled-sol[5]	Z-score[4]	Solubility[6]	Z-score[4]	
Group 3	Putative endolysin 177	249	0.569	-0.269	0.680	5.559	0.186	-1.638	3.652
Group 1	Putative endolysin 117	481	0.735	1.114	0.354	-0.276	0.327	0.903	1.741
Group 1	Putative endolysin 129	481	0.776	1.465	0.343	-0.473	0.311	0.610	1.601
Group 1	Putative endolysin 121	481	0.749	1.239	0.332	-0.670	0.330	0.951	1.520
Group 1	Putative endolysin 159	481	0.718	0.973	0.332	-0.670	0.343	1.182	1.485
Group 1	Putative endolysin 132	481	0.750	1.241	0.343	-0.473	0.312	0.631	1.398
Group 1	Putative endolysin 147	481	0.727	1.052	0.332	-0.670	0.332	0.981	1.364
Group 1	Putative endolysin 143	481	0.741	1.167	0.322	-0.849	0.334	1.015	1.333
Group 1	Putative endolysin 174	481	0.722	1.006	0.332	-0.670	0.332	0.986	1.322
Group 1	Putative endolysin 150	481	0.732	1.092	0.322	-0.849	0.336	1.063	1.306

[1] The top 10 of the 114 putative endolysins are presented in this table, and the complete dataset for predicted solubility is presented in Table S3.
[2] Group 1, CHAP domain + Ami_2 domain + SH3_5 domain; group 2, CHAP domain + Ami_2 domain + SH3 superfamily domain; group 3, Amidase_3 domain + SH3b domain.
[3] Solubility, predicted solubility scores based on Soluprot.
[4] Z-score = $(x - \mu)/\sigma$, where x is the raw score, μ is the mean, and σ is the standard deviation.
[5] Scaled-sol, predicted solubility scores based on Protein-Sol.
[6] Solubility, predicted solubility scores based on SKADE.
[7] Sum, the sum of Z-scores calculated from Soluprot, Protein-Sol, and SKADE.

2.4. Overexpression and structural analysis of selected endolysins

Of the two endolysin candidates, only recombinant putative endolysin 117 was successfully overexpressed and purified in its soluble form, and we renamed it LyJH1892. LyJH1892 consists of 481 amino acids, and its molecular weight was 53.8 kDa. The major band of purified soluble LyJH1892 was similar to the expected molecular mass (53.8 kDa) upon SDS-PAGE (Figure 4a). The conserved domains of LyJH1892 consisted of three distinct domain architectures (Figure 4b); (1) N-terminal CHAP domain (Accession: pfam05257, e-value = 2.21×10^{-8}, and bit-score = 51.3), (2) Amidase_2 domain (Accession: smart00644, e-value = 5.22×10^{-24}, and bit-score = 96.7), and (3) SH3_5 domain (Accession: pfam08460, e-value = 6.20×10^{-22}, and bit-score = 88.9) (Table S2). The 3D structure of LyJH1892 was predicted using AlphaFold2 and visualized in ribbon form and Connolly surface form in ChimeraX 1.3 (Figure 4c and d). The overall secondary structure of LyJH1892 consisted of an alpha-helix (22.7%), beta-sheet (21.8%), and random coil (55.5%). The CHAP domain showed 27.7% alpha helices, 22.9% beta-sheets, and 49.4% random coils. Furthermore, the Amidase_2 domain showed a secondary structure composition similar with the CHAP domain (Amidase_2 domain; alpha helix: 21.6%, beta-sheet: 18.4%, and random coil: 60.0%). Unlike both EADs, the SH3_5 domain consisted of 51.5% beta sheets and 48.8% random coils.

 doi:10.20944/preprints202302.0072.v1

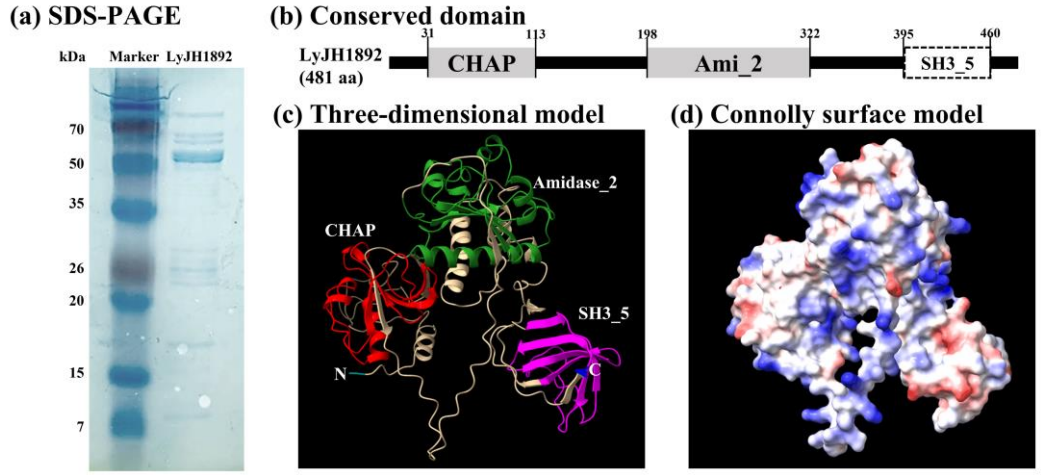

Figure 4. Structural characteristics of endolysin LyJH1892. (a) Analysis of endolysin LyJH1892 by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Lane 1 contained a protein molecular weight marker and lane 2 contained purified LyJH1892. (b) The conserved domain of LyJH1892. The gray square represents the enzymatically active domain, and the white square describes the cell wall binding domain. (c) Three-dimensional model of LyJH1892 predicted by AlphaFold2 on Colab Fold notebook. (d) Connolly surface of LyJH1892 created by ChimeraX 1.3. Blue to red color represents most positive to most negative polar activities.

2.5. Characterization of Recombinant LyJH1892

We assessed the lytic activity of LyJH1892 against *S. aureus* (NCCP 16830) at various pH values, temperatures, NaCl concentrations, and metal ions to identify the optimal conditions for LyJH1892. The lytic activity of LyJH1892 was significantly higher at pH 6.0 and 9.0 than at 10.0 (P = 0.0089). Additionally, it was highly stable at pH 6.0–9.0 (Figure 5a). The optimal temperature for LyJH1892 activity was 25 °C, and the moderate activity remained at 4–37 °C (above 50%); however, the lytic activity of LyJH1892 drastically decreased above 50 °C (Figure 5b, P = 0.0065). The highest lytic activity of LyJH1892 was detected at 0 mM of NaCl (Figure 5c, P = 0.0054), and the moderate activity was maintained at 0–62.5 mM of NaCl (above 60%), whereas the lytic activity decreased above 125 mM of NaCl (approximately 30% or less). In the metal ion-dependent test, the addition of 5 mM EDTA to LyJH1892 significantly decreased its lytic activity compared to pure LyJH1892 (None) (Figure 5d, P = 0.0065). The addition of 10 mM Ca^{2+}, Mg^{2+}, and Mn^{2+} to EDTA-treated LyJH1892 restored lytic activity to > 65%. Nonetheless, it did not fully recover the lytic activity of LyJH1892 (Figure 5d).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0072.v1

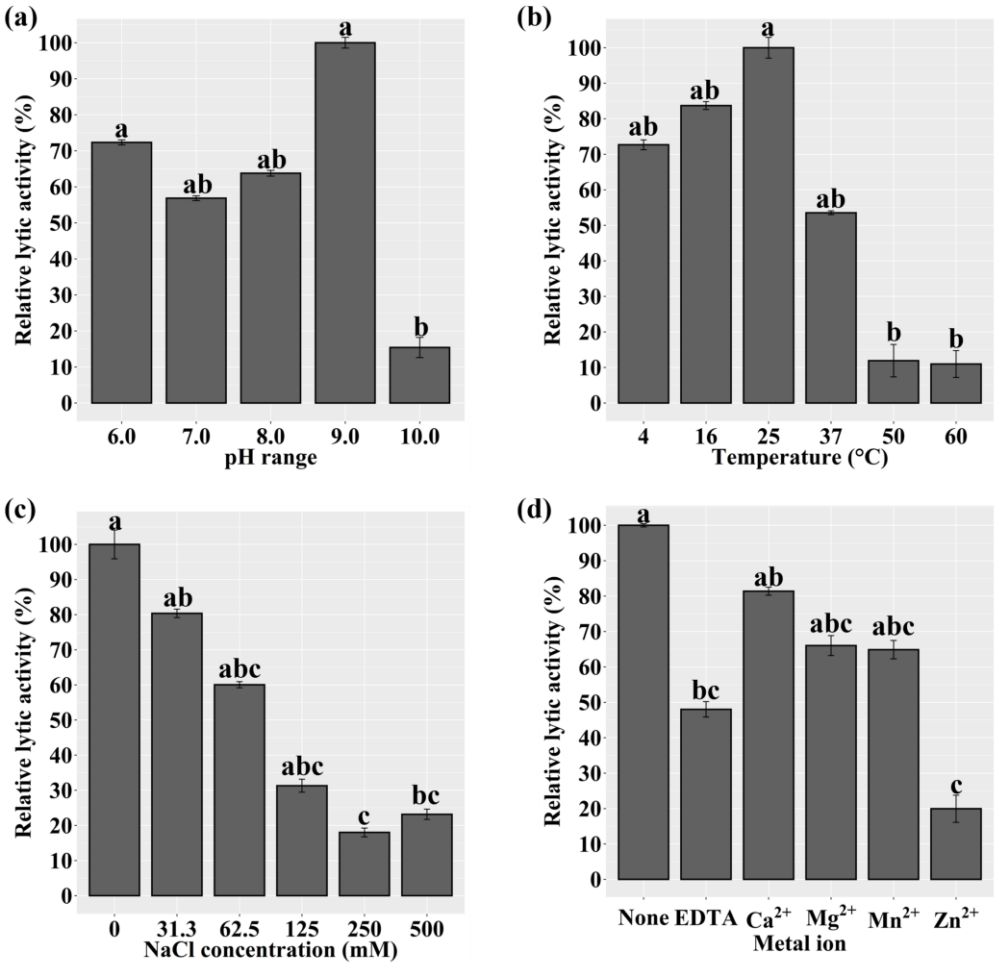

Figure 5. Lytic activity of LyJH1892 at various (a) pHs, (b) temperatures, (c) NaCl concentrations, and (d) metal ions against *Staphylococcus aureus* (NCCP 16830). Data are shown as means ± standard deviation of triplicate assays. In Figure 5d, None, pure LyJH1892; EDTA, purified LyJH1892 after incubating with 5 mM ethylenediaminetetraacetic acid (EDTA); Other metal cation treatments (Ca^{2+}, Mg^{2+}, Mn^{2+}, and Zn^{2+}), LyJH1892 after adding 10 mM metal ions to EDTA treated LyJH1892.

2.6. Optimal lytic activity and lytic spectrum of LyJH1892

Under optimal conditions (pH 9.0, 25 °C, and no addition of NaCl or metal ions), the dose-response test of LyJH1892 was evaluated against *S. aureus* (NCCP16830). The lytic activity of LyJH1892 was increased in a dose-dependent manner, and the addition of the highest dosage (100 μg/ml concentration) lysed approximately 85% of *S. aureus* (Figure 6a). To evaluate the lytic spectrum of LyJH1892, we further analyzed the lytic activity test under optimal conditions against two strains of normal *S. aureus* (NCCP16830 and KVCC-BA0500624), two strains of MRSA (NCCP14754 and NCCP14567), and seven species of coagulase-negative Staphylococci (CNS) (*S. epidermidis* [KVCC-BA0001452], *S. chromogenes* [KCTC3579], *S. hyicus* [KVCC-BA0001860], *S. haemolyticus* [NCCP14693], *S. warneri* [NCCP16234], *S. simulans* [NCCP16236], and *S. xylosus* [KCCM40887]), and *S. agalactiae* (NCCP14729) (Table 2). Among the *S. aureus* strains, LyJH1892 potently lysed both normal *S. aureus* (Figure 6b, relative lytic activity > 95%). Interestingly, LyJH1892 exhibited more potent lytic activity against MRSA than normal *S. aureus* strains (Figure 6b, relative lytic activity > 125%). Among the CNS species, the highest lytic activity was observed against *S. epidermidis* (relative lytic activity, approximately 145%). Additionally, the lytic activity was highly retained against *S. chromogenes*, *S. hyicus*, and *S. haemolyticus* (relative lytic

 doi:10.20944/preprints202302.0072.v1

activity > 70%) (Figure 6b). However, LyJH1892 showed low lytic activity against two CNS species (*S. warneri* and *S. simulans*) and *S. agalactiae* (Figure 6b).

Table 2. Bacterial strains and growth condition

Bacterial strains	Purpose	Growth condition[1]
Staphylococcus aureus (NCCP 16830)	Indicator strain	BHI broth
Escherichia coli DH5α	Cloning host	LB broth
E. coli BL21(DE3)	Expression host	LB broth
S. aureus (KVCC-BA0500624)	Lytic spectrum	BHI broth
MRSA (NCCP 14567)	Lytic spectrum	BHI broth
MRSA (NCCP 14754)	Lytic spectrum	BHI broth
Staphylococcus hyicus (KVCC-BA0001860)	Lytic spectrum	BHI broth
Staphylococcus epidermidis (KVCC-BA0001452)	Lytic spectrum	BHI broth
Staphylococcus haemolyticus (NCCP 14693)	Lytic spectrum	BHI broth
Staphylococcus simulans (NCCP 16236)	Lytic spectrum	BHI broth
Staphylococcus warneri (NCCP 16234)	Lytic spectrum	BHI broth
Staphylococcus chromogenes (KCTC 3579)	Lytic spectrum	BHI broth
Staphylococcus xylosus (KCCM 40887)	Lytic spectrum	BHI broth
Streptococcus agalactiae (NCCP 14729)	Lytic spectrum	BHI broth

MRSA, methicillin-resistant *Staphylococcus aureus*; BHI, brain heart infusion; LB, Luria-Bertani.
[1] All bacterial strains used in this study were grown aerobically at 37 °C.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2023 doi:10.20944/preprints202302.0072.v1

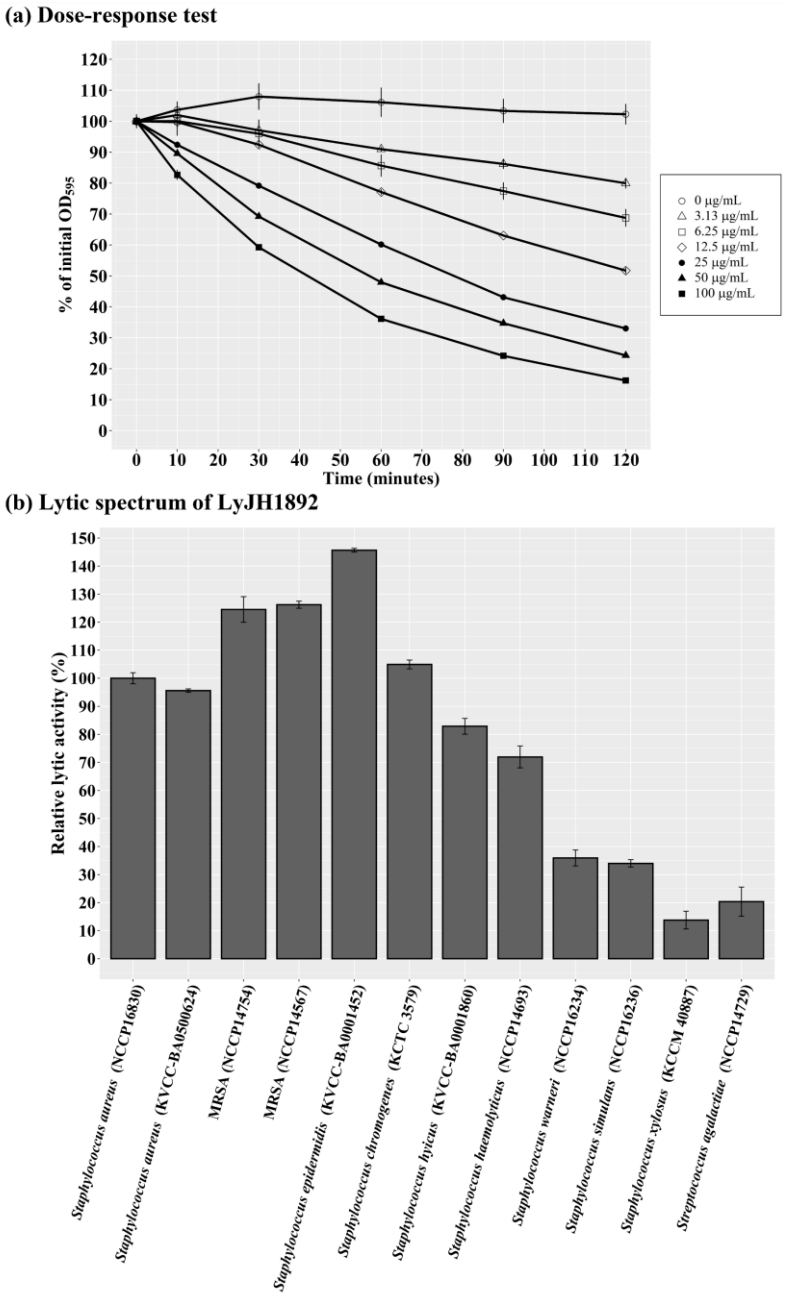

Figure 6. Optimal lytic activity and lytic spectrum of LyJH1892. (a) Dose-response test against *Staphylococcus aureus* (NCCP 16830). (b) The lytic spectrum of LyJH1892. All lytic tests were performed under optimal conditions of LyJH1892 (pH 9.0, 25℃, and no addition of NaCl and metal ions).

3. Discussion

In this study, we introduced a logical strategy to develop putative endolysins targeting MRSA based on genome information and bioinformatics tools instead of the traditional bacteriophage screening approach. We identified 272 putative endolysin candidates from 474 MRSA genome sequences and successfully developed and characterized a new endolysin, LyJH1892, which can kill *S. aureus* and MRSA.

Biosynthesis of the peptidoglycan layer, which is composed of disaccharide *N*-acetylglucosamine-*N*-acetylmuramic acid with peptide stems, is a fundamental process for bacterial survival during growth and division [22]. The crosslinking reaction between *N*-acetylmuramic acid and peptide stems is mediated by transpeptidases, also known as PBPs [23]. In general, native PBPs in *S. aureus* include PBP1, PBP2, PBP3, and PBP4 [24];

 doi:10.20944/preprints202302.0072.v1

however, MRSA possesses a unique PBP called PBP2a [25,26]. β-lactam antibiotics bind to native PBPs of *S. aureus* and block the reconstruction of the peptidoglycan layer, while MRSA can survive in the presence of β-lactam antibiotics because PBP2a has a low affinity for β-lactam antibiotics [24]. A previous study suggested that the *mec*A gene, which encodes PBP2a, could be a useful marker for methicillin resistance in *S. aureus* [27]. Therefore, we used PBP2a as a computational marker to identify MRSA among the whole genome of *S. aureus* listed in the NCBI database. In this study, we found that approximately 67% of *S. aureus* was detected as MRSA among 707 whole genomes. A meta-analysis including 37 studies reported that the pooled prevalence of MRSA among *S. aureus* in China was 15% [28]. In Europe, the proportion of MRSA among *S. aureus* was 19% in 2008 [29]. The prevalence of MRSA in our study was significantly higher than that reported in previous studies. However, several studies have reported a high prevalence of MRSA with similar or even higher prevalence than that in our study: 72 % of MRSA in Eritrea [30], 80% in Peru [31], and 90% in Colombia [32]. According to the existing literature on the prevalence of MRSA among *S. aureus* strains, several factors are related to the significant variation (e.g., geographical variation, study design, MRSA detection methods, and study population) [28,33]. For instance, several studies suggested that there is a clear geographical variation in the prevalence of MRSA [28,33]. Another possible explanation for the high prevalence of MRSA is the increased detection rate, which results in a biased proportion of *S. aureus* genome sequences in the NCBI database.

In the present study, we identified 272 putative endolysin sequences from MRSA genomes, showing variations in amino acid sequences and conserved domains. The CHAP domain is known as an EAD with two different peptidoglycan catalytic activities: *N*-acetylmuramoyl-L-alanine amidase and endopeptidase splitting cross-linked peptide stems, especially D-alanine-glycine [34]. In addition, a previous study suggested that the CHAP domain is predominantly found in phage endolysins that infect gram-positive bacteria, especially *Staphylococcus* spp. [35]. SH3 domains were first identified in eukaryotes and are generally involved in cell-cell communication and intracellular signaling from the cell membrane to the nucleus by binding to proline-rich polypeptides [36,37]. However, hundreds of SH3 domains, which are composed of SH3b, SH3-3, or SH3-5 types, were also found in bacteria and bacteriophage proteins, such as autolysin and endolysin structures [35,38]. Many studies have verified the role of bacterial SH3 domains as binding modules found in endolysins targeting *S. aureus* [39], several *Bacillus* species [40], and *Listeria monocytogenes* [41]. In addition, host specificity or host range of endolysins is determined by the type of C-terminal CBD [10,42,43]. Therefore, we only focused on the 114 putative endolysins (approximately 72.6%) with multi-conserved domain combinations possessing both EAD and CBD domains. In the present study, we found that 53.5% of endolysins had a CHAP domain as an EAD, and all endolysins exhibited SH3-like domains as CBD. In a previous study, Becker, Foster-Frey, Stodola, Anacker and Donovan [37], who investigated *Staphylococcus* phage proteins containing SH3 domains (SH3b and SH3_5), reported that SH3 containing endolysins were divided into five groups based on their sequence similarity (>90%) (group 1, CHAP – Amidase_3 and SH3b domain; group 2 and 3, CHAP, Amidase_2, and SH3b domains; group 4, CHAP and SH3b domains; and group 5, glycyl-glycine endopeptidase and SH3b domain). Recently, Chang and Ryu [18] reported that 98 *Staphylococcus* endolysins, which were identified from 98 *Staphylococcus* phage sequences enrolled in NCBI, were divided into six groups based on their conserved domain architecture (group 1, CHAP + unknown CBD ; group 2, CHAP + SH3 domain; group 3, CHAP + Amidase_2 + and SH3 domain; group 4, CHAP, Amidase_3, and SH3 domains; group 5, CHAP, Amidase_3, and CBD (unknown); and group 6, Peptidase_M23, Amidase_2, and SH3 domains). The majority of endolysins from the phage source represented a combination of CHAP, Amidase, and SH3-like domains (Becker, Foster-Frey, Stodola, Anacker and Donovan [37]], 89.5% and Chang and Ryu [18], 68.4%). Moreover, there were few endolysins that did not have a CHAP domain (Becker, Foster-Frey, Stodola, Anacker and Donovan [37]], 3.5% and Chang and Ryu [18]], 1.0%). Interestingly, in the present study,

 doi:10.20944/preprints202302.0072.v1

a different type of domain architecture (group 3) was observed, occupying 46.5% of total putative endolysins, which consisted of Amidase_3 and SH3b domains. Based on the amino acid sequences, group 3 showed the highest sequence similarity (Amidase_3 domain, 95.1%; and SH3_b domain, 88.4%). In this study, the overall sequence similarity of groups 1 and 2 was highly conserved, and that of CHAP domains from groups 1 and 2 exhibited high conservation (overall CHAP domains, 82.4%). However, Becker, Foster-Frey, Stodola, Anacker and Donovan [37] reported that CHAP domains within groups showed high conservation (>90% identity), although decreased in comparison to overall CHAP domains (less than 50% identity). In addition, they also reported a relatively higher sequence similarity between Amidase_2 and Amidase_3 (53% identity) than that in our study (overall sequence similarity in amidase-like domains, 12.7%). In the present study, we found that the SH3-like domains consisted of 61–68 amino acids and lower sequence similarities with 10 perfectly conserved amino acid positions (glycine, 3; tyrosine, 3; valine, 1; arginine, 1; cysteine, 1; and tryptophan, 1). However, Becker, Foster-Frey, Stodola, Anacker and Donovan [37] reported that the alignment of the 11 SH3_5 domains from endolysins identified from *Staphylococcus* phages showed six perfectly conserved amino acid positions (glycine, 2; glutamic acid, 1; arginine, 1; tyrosine, 1; tryptophan, 1). The endolysins identified in this study showed low sequence similarity and conserved amino acid positions mainly in SH3-like domains with endolysins from *Staphylococcus* phages. Therefore, we speculated that the newly identified 114 putative endolysins derive from different ancestors with endolysins described in previous studies.

The production of heterologous proteins in *E. coli* systems often has problems such as low yield in a soluble form or production of insoluble proteins [44]. Although several empirical processes can be used to optimize the production of heterologous proteins in soluble form (*e.g.*, designing promoters, codon optimization, and changing culture conditions) [45], these processes are generally laborious and time-consuming. There are many bioinformatic tools to predict protein solubility using sequence-based and machine learning-based methods [46,47]. In the present study, we evaluated the predicted protein solubility using three different bioinformatic tools using a sequence-based method (ProteinSol) [48] and machine-learning-based methods (SoluProt and SKADE) [49,50]. In the present study, protein solubilities from machine learning-based methods were generally higher in putative endolysins belonging to group 1 than in group 3. Conversely, opposite trends were observed in the sequence-based method. Based on the integrated protein solubility indices, we attempted to overexpress heterologous endolysins (putative endolysins 177 and 117). Despite the high solubility scores, only putative endolysin 117 (LyJH1892) was successfully produced in its soluble form. Overall, we propose that the use of solubility prediction tools based on machine learning methods could be a useful strategy to identify endolysins among various putative candidates. However, due to limited data, further research should be conducted to verify more accurate standards for the selection of endolysin candidates using solubility prediction tools.

The increasing appearance of livestock-associated MRSA (LA-MRSA) has become a serious problem because this major pathogen causes bovine mastitis and lethal infections to livestock-related workers [51,52]. In addition, CNS are considered an emerging risk factor for bovine mastitis, inducing a negative effect on udder health and intramammary infections [7,53-56]. In this study, we successfully characterized endolysin LyJH1892 with lytic activity against *S. aureus* at different pH, temperature, and NaCl concentrations. Notably, the lytic activity of LyJH1892 was higher against MRSA than against *S. aureus.* This suggests that our strategy is useful in the development of specific endolysins against antibiotic-resistant strains. In general, the peptidoglycan of Staphylococci belongs to the peptidoglycan subgroup A3α, which is cross-linked by pentaglycine bridges [57]. Lysostaphin is a glycyl-glycine endopeptidase secreted by *S. simulans* with a known bactericidal effect against *S. aureus* [14]. Similar to lysostaphin, ALE-1 is a glycyl-glycine endopeptidase secreted by *Staphylococcus capitis* EPK1, which also has lytic activity against *S. aureus* and other *Staphylococcus* species [13,15]. Both peptidoglycan hydrolases are known

 doi:10.20944/preprints202302.0072.v1

to have SH3-like domains (SH3b domain) that share a high degree of homology. Additionally, previous studies have shown that the SH3b domains of these enzymes require intact penta-glycine cross-bridges for binding [58,59]. However, LyJH1892 showed potent killing activity against only four of the seven CNS species, including *S. chromogenes*, *S. haemolyticus*, *S. epidermidis*, and *S. hyicus*. These unexpected results from the specificity spectrum have several possible reasons. Several previous studies have demonstrated that truncated endolysins consisting of only EADs also lyse Staphylococcal species, indicating that the binding spectrum of endolysin is not determined by the characteristics of CBD alone [60,61]. In addition, Lu, Fujiwara, Komatsuzawa, Sugai and Sakon [58] reported that the outside region of the SH3b domain in ALE-1 is essential for the full recovery of binding affinity. Furthermore, Lu, Fujiwara, Komatsuzawa, Sugai and Sakon [58] proved that the composition of penta-amino acids in the peptidoglycan layer also affected the lytic spectrum of the SH3b domain of ALE-1, especially showing a strong preference for penta-glycine. Furthermore, there are variations in interpeptide bridges among Staphylococci and Streptococci belonging to peptidoglycan subgroup A3α [57,62]. This could be one of the reasons for the unexpected result from the present lytic spectrum. However, based on the potent lytic activity of LyJH1892 on MRSA and the important pathogenic CNS, LyJH1892 could be a useful biocontrol agent for bovine mastitis.

4. Materials and Methods

4.1. Bacterial strains and growth conditions

The bacterial strains used in the present study are listed in Table 2. *Escherichia coli* cloning strain DH5α and expression strain BL21 (DE3) were aerobically grown at 37 °C in Luria-Bertani (LB) broth. *S. aureus* (NCCP 16830), MRSA (NCCP 14567), MRSA (NCCP14754), *Staphylococcus haemolyticus* (NCCP 14693), *Staphylococcus simulans* (NCCP 16236), *Staphylococcus warneri* (NCCP 16234), and *Streptococcus agalactiae* (NCCP 14729) were obtained from the National Culture Collection for Pathogens (Cheongju, Korea). *S. aureus* (KVCC-BA0500624), *Staphylococcus hyicus* (KVCC-BA0001860), and *Staphylococcus epidermidis* (KVCC-BA0001452) were obtained from the Korea Veterinary Culture Collection (Gimcheon, Korea). *Staphylococcus chromogenes* (KCTC 3579) and *Staphylococcus xylosus* (KCCM 40887) were obtained from the Korea Collection for Type Cultures (Jeongeup, Korea) and the Korean Culture Center of Microorganisms (Seoul, Korea), respectively. All Staphylococcal species and *S. agalactiae* were grown aerobically at 37 °C in brain heart infusion broth.

4.2. Collection of putative endolysins against MRSA

Complete *S. aureus* genome sequences were retrieved from the National Center for Biotechnology Information (NCBI) (retrieval date: 06.02.2020) and annotated using a Rapid Annotation using Subsystems Technology server [63]. After annotation, *S. aureus* containing PBP2a was considered MRSA [24] and putative endolysin sequences were retrieved from their genomes.

4.3. Conserved domain analysis, sequence alignment, and solubility prediction of putative endolysins

The conserved domains of putative endolysins were analyzed using the NCBI conserved domain database [64], and the distribution of the conserved domains was visualized using the ggplot2 packages [65] in the R software (version 4.1.3) [66]. Putative endolysins containing both EAD and CBD were selected. Multisequence-alignments based on amino acid sequences were carried out using the E-INS-i algorithm implemented in multiple alignment using fast Fourier transform (MAFFT v7.505) [67]. Alignments were used to perform phylogenetic reconstructions of putative endolysins using FastTree [68] and default parameters (amino acid distances BLOSUM45 and the Jones-Taylor-Thorton

 doi:10.20944/preprints202302.0072.v1

model), and the phylogenetic tree was visualized using ggtree [69] in the R software (version 4.1.3) [66]. Among putative endolysin sequences with CBD, protein solubility was predicted based on various prediction tools, including SoluProt [49], Protein-sol [48], and SKADE [50]. The solubility value was then transformed into a Z-score, followed by a selection of endolysin candidates based on the top two of Z-score sums.

4.4. Identification, cloning, and overexpression of endolysin LyJH1892

Based on the solubility index, putative endolysin 177 and 117 were selected as endolysin candidates and named LyJH1508 and LyJH1892, respectively. The sequences of LyJH1508 and LyJH1892 were acquired from the whole genome sequence of *S. aureus* (GenBank accession numbers: GCA_003193885.1 and GCA_000568455.1, respectively). Both endolysins were chemically synthesized after codon optimization to facilitate the overexpression of the recombinant protein in *E. coli*, and the chemically synthesized endolysin genes were amplified by polymerase chain reaction (PCR) using HiPi™ plus thermostable Taq DNA polymerase (Elpis-biotech, Daejeon, Korea). Purified DNA fragments were digested using the restriction enzymes *Bam*H1 and *Xho*1 (New England Biolabs Inc., Ipswich, MA, USA) and then cloned into the expression vector pET28b (Novagen Inc., Madison, WI, USA) with an N-terminal hexa histidine-tag (6xHis tag) sequence. The cloned pET28b vector was transformed into competent *E. coli* BL21 (DE3) cells. The transformants were grown in LB broth until the optical density at 595 nm (OD_{595nm}) reached 0.4, and 0.5 mM isopropyl-β-D- thiogalactopyranoside (IPTG) was added to the medium to overexpress the target recombinant endolysin. Cells were further incubated for 18 h at 16 °C. Harvested cells were resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and the resuspended cells were lysed on ice using a sonicator (KYY-80, Korea Process Technology Co., Ltd., Seoul, Korea). After centrifugation at 12,000 × g for 15 min, the crude extracts were purified using Nuvia™ IMAC resin charged with Ni (Bio-Rad Laboratories, Inc., CA, USA), according to the manufacturer's instructions, followed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. The purified endolysin was pooled and dialyzed against the elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, and pH 8.0).

4.5. Structure prediction of LyJH1892

To predict the three-dimensional (3D) structure of LyJH1892, we used Colab Fold notebook, which predicts protein structure using AlphaFold2 combined with a fast multiple sequence alignment generation stage using MMseq2 [70]. The predicted structure was visualized using the ChimeraX 1.3 [71].

4.6. Characterization and lytic spectrum of LyJH1892

S. aureus (NCCP 16830) was used as the reference strain for the lytic test of LyJH1892. *S. aureus* (NCCP 16830) was grown to an OD_{595nm} of 0.8–1.0 and then harvested. To characterize the endolysin activity, we tested the lytic activity of LyJH1892 under various conditions (pH, temperature, NaCl concentration, and metal ions), and 20 μL of LyJH1892 (100 μg/mL) was added to 96-well plates (SPL Life Sciences Co., Ltd., Pocheon, Korea) containing cell suspensions (180 μL). For control samples, 20 μL of elution buffer was added instead of endolysin. The OD_{595nm} values were monitored using an iMark microplate reader (Bio-Rad Laboratories Inc., Hercules, CA, USA). The lytic activity of LyJH1892 was calculated after 2 h as follows: $100 \times (\Delta OD_{595}$ treated with LyJH1892 $- \Delta OD_{595}$ treated with elution buffer)/initial OD_{595}. To determine the optimal pH range for endolysin activity, the indicator strain was resuspended in 50 mM sodium phosphate (pH 6.0, 7.0, and 8.0) and sodium glycine buffer (pH 9.0, 10.0), and the mixture was incubated at 25 °C for 120 min. To determine the optimal temperature of LyJH1892, the mixture was incubated at different temperatures (4, 16, 25, 37, 50, and 60 °C) for 120 min. The effect of NaCl on LyJH1892 activity was evaluated by adding 0, 31.3, 62.5, 125, 250, and 500 mM

 doi:10.20944/preprints202302.0072.v1

NaCl to the empirically determined pH buffer. The effect of divalent cations on LyJH1892 was measured as described by Kim, Lee, Kwon and Seo [21]. Briefly, LyJH1892 (100 μg/mL) was incubated with 5 mM ethylenediaminetetraacetic acid (EDTA) at 25°C for 30 min to chelate divalent cations attached to the endolysin, and EDTA was removed by changing the elution buffer to Amicon Ultra-4 (10 kDa) (Merck KGaA, Darmstadt, Germany) [72]. LyJH1892 was incubated with 10 mM $CaCl_2$, $MgCl_2$, $MnCl_2$, and $ZnCl_2$ at 25 °C for 30 min, and the lytic activity of LyJH1892 treated with EDTA and various divalent cations was assessed. To determine the dose-dependent response of LyJH1892 under optimal conditions, the indicator strain was resuspended in an empirically determined buffer (sodium glycine buffer, pH 9.0, and NaCl 0 mM), and serially diluted LyJH1892 (20 μL, 3.125-100 μg/mL concentration) and elution buffer (20 μL, for control) were added to the suspension. The mixture was then incubated at 25 °C, and the OD_{595nm} values were monitored at 0, 10, 30, 60, 90, and 120 min. A list of the bacterial strains used in the lytic spectrum test is presented in Table 2. All bacterial species used in the lytic spectrum test were grown as described above to an OD_{595nm} 0.8–1.0, harvested and resuspended in the empirically determined buffer. The lytic activity of LyJH1892 was evaluated after incubation for 2 h at 25 °C. All experiments were conducted in triplicate.

4.7. Statistical analysis

Statistical analysis was performed using the R software (version 4.1.3) [66]. The data derived from the characterization of the LyJH1892 lytic activity were analyzed using the non-parametric Kruskal-Wallis test using the Kruskal-Wallis test function because the residuals did not meet the normal distribution. A post-hoc Dunn test (Dunn Test function in the FSA package) [73] was used to compare differences among treatments, if a significant difference was observed. All *P*-values were adjusted using the Benjamini-Hochberg false discovery rate, and statistical significance was set at $P < 0.05$.

5. Conclusions

In summary, we present a novel strategy to identify new endolysin candidates against MRSA from prophage sources by combining bioinformatics approaches. A total of 272 putative endolysin candidates were detected from the MRSA genome sequences, and only 114 putative endolysins exhibited multi-conserved domain combinations possessing EAD and CBD domains. In addition, we successfully developed a new endolysin, LyJH1892, which has broad lytic activity against *S. aureus*, MRSA, and several CNS, suggesting that LyJH1892 could be a useful biocontrol agent for bovine mastitis. These findings provide a rapid and useful strategy for the development of specific endolysins against antibiotic-resistant bacterial strains.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Table S1: List of collected genome of Staphylococcus aureus from National Center for Biotechnology Information (collecting date: 06.02.2020) and Penicillin-binding protein 2a.; Table S2: Information of conserved domain analysis for identified putative endolysins.; Table S3: Predicted recombinant protein solubility of putative endolysins based on Soluprot, Protein-Sol, and SKADE.; Figure S1: Multi-alignment of putative endolysins belonging to group 1 based on amino acids. (a) Multi-alignment of the cysteine, histidine-dependent amidohydrolases/peptidases domain (Accession: pfam05257 and sequence identity: 83.5%). (b) Multi-alignment of the Ami_2 domain (Accession: smart00644 and sequence identity: 82.4%). (c) Multi-alignment of the SH3_5 domain (Accession: pfam08460 and sequence identity: 75.8%).; Figure S2: Multi-alignment of putative endolysins belonging to group 3 based on amino acids. (a) Multi-alignment of the Ami_3 domain (Accession: pfam01520 and sequence identity: 95.1%). (b) Multi-alignment of the SH3_b domain (Accession: smart00287 and sequence identity: 88.4%).

Author Contributions: Conceptualization, H.K. and J.S.; methodology, H.K.; software, H.K.; validation, J.S.; formal analysis, H.K.; investigation, H.K. and J.S.; resources, H.K.; data curation, J.S.; writing—original draft preparation, H.K.; writing—review and editing, J.S.; visualization, H.K.;

supervision, J.S.; project administration, J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education(2020R1A6A3A13070583).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Haddad Kashani, H.; Schmelcher, M.; Sabzalipoor, H.; Seyed Hosseini, E.; Moniri, R. Recombinant endolysins as potential therapeutics against antibiotic-resistant Staphylococcus aureus: current status of research and novel delivery strategies. *Clinical microbiology reviews* **2018**, *31*, e00071-00017. https://doi.org/10.1128/CMR.00071-17.
2. Kluytmans, J.; Van Belkum, A.; Verbrugh, H. Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. *Clinical microbiology reviews* **1997**, *10*, 505-520. https://doi.org/10.1128/CMR.10.3.505.
3. Gutiérrez, D.; Fernández, L.; Rodríguez, A.; García, P. Are phage lytic proteins the secret weapon to kill Staphylococcus aureus? *MBio* **2018**, *9*, e01923-01917. https://doi.org/10.1128/mBio.01923-17.
4. Wertheim, H.F.L.; Melles, D.C.; Vos, M.C.; van Leeuwen, W.; van Belkum, A.; Verbrugh, H.A.; Nouwen, J.L. The role of nasal carriage in Staphylococcus aureus infections. *The Lancet Infectious Diseases* **2005**, *5*, 751-762, https://doi.org/10.1016/S1473-3099(05)70295-4. https://doi.org/10.1016/S1473-3099(05)70295-4.
5. Foster, T.J. Antibiotic resistance in Staphylococcus aureus. Current status and future prospects. *FEMS microbiology reviews* **2017**, *41*, 430-449. https://doi.org/10.1093/femsre/fux007.
6. Sakoulas, G.; Moellering, R.C., Jr. Increasing Antibiotic Resistance among Methicillin-Resistant Staphylococcus aureus Strains. *Clinical Infectious Diseases* **2008**, *46*, S360-S367, https://doi.org/10.1086/533592. https://doi.org/10.1086/533592.
7. Cheng, W.N.; Han, S.G. Bovine mastitis: Risk factors, therapeutic strategies, and alternative treatments—A review. *Asian-Australasian journal of animal sciences* **2020**, *33*, 1699. https://doi.org/10.5713/ajas.20.0156.
8. Guimarães, F.; Manzi, M.; Joaquim, S.; Richini-Pereira, V.; Langoni, H. Outbreak of methicillin-resistant Staphylococcus aureus (MRSA)-associated mastitis in a closed dairy herd. *Journal of Dairy Science* **2017**, *100*, 726-730. https://doi.org/10.3168/jds.2016-11700.
9. Juhász-Kaszanyitzky, É.; Jánosi, S.; Somogyi, P.; Dán, Á.; vanderGraaf van Bloois, L.; Van Duijkeren, E.; Wagenaar, J.A. MRSA transmission between cows and humans. *Emerging infectious diseases* **2007**, *13*, 630. https://doi.org/10.3201/eid1304.060833.
10. Schmelcher, M.; Donovan, D.M.; Loessner, M.J. Bacteriophage endolysins as novel antimicrobials. *Future microbiology* **2012**, *7*, 1147-1171. https://doi.org/10.2217/fmb.12.97.
11. Donovan, D.M. Bacteriophage and peptidoglycan degrading enzymes with antimicrobial applications. *Recent patents on biotechnology* **2007**, *1*, 113-122. https://doi.org/10.2174/187220807780809463.
12. Love, M.J.; Bhandari, D.; Dobson, R.C.; Billington, C. Potential for bacteriophage endolysins to supplement or replace antibiotics in food production and clinical care. *Antibiotics* **2018**, *7*, 17. https://doi.org/10.3390/antibiotics7010017.
13. Cisani, G.; Varaldo, P.E.; Grazi, G.; Soro, O. High-level potentiation of lysostaphin anti-staphylococcal activity by lysozyme. *Antimicrobial agents and chemotherapy* **1982**, *21*, 531-535. https://doi.org/10.1128/AAC.21.4.531.

14. Kumar, J.K. Lysostaphin: an antistaphylococcal agent. *Applied microbiology and biotechnology* **2008**, *80*, 555-561. https://doi.org/10.1007/s00253-008-1579-y.
15. Sugai, M.; Fujiwara, T.; Akiyama, T.; Ohara, M.; Komatsuzawa, H.; Inoue, S.; Suginaka, H. Purification and molecular characterization of glycylglycine endopeptidase produced by Staphylococcus capitis EPK1. *Journal of bacteriology* **1997**, *179*, 1193-1202. https://doi.org/10.1128/jb.179.4.1193-1202.1997.
16. Becker, S.C.; Foster-Frey, J.; Donovan, D.M. The phage K lytic enzyme LysK and lysostaphin act synergistically to kill MRSA. *FEMS microbiology letters* **2008**, *287*, 185-191. https://doi.org/10.1111/j.1574-6968.2008.01308.x.
17. Sundarrajan, S.; Raghupatil, J.; Vipra, A.; Narasimhaswamy, N.; Saravanan, S.; Appaiah, C.; Poonacha, N.; Desai, S.; Nair, S.; Bhatt, R.N. Bacteriophage-derived CHAP domain protein, P128, kills Staphylococcus cells by cleaving interpeptide cross-bridge of peptidoglycan. *Microbiology* **2014**, *160*, 2157-2169. https://doi.org/10.1099/mic.0.079111-0.
18. Chang, Y.; Ryu, S. Characterization of a novel cell wall binding domain-containing Staphylococcus aureus endolysin LysSA97. *Applied microbiology and biotechnology* **2017**, *101*, 147-158. https://doi.org/10.1007/s00253-016-7747-6.
19. Altermann, E.; Schofield, L.R.; Ronimus, R.S.; Beattie, A.K.; Reilly, K. Inhibition of rumen methanogens by a novel archaeal lytic enzyme displayed on tailored bionanoparticles. *Frontiers in microbiology* **2018**, 2378. https://doi.org/10.3389/fmicb.2018.02378.
20. Swift, S.M.; Waters, J.J.; Rowley, D.T.; Oakley, B.B.; Donovan, D.M. Characterization of two glycosyl hydrolases, putative prophage endolysins, that target Clostridium perfringens. *FEMS microbiology letters* **2018**, *365*, fny179. https://doi.org/10.1093/femsle/fny179.
21. Kim, H.; Lee, H.G.; Kwon, I.; Seo, J. Characterization of endolysin LyJH307 with antimicrobial activity against Streptococcus bovis. *Animals* **2020**, *10*, 963. https://doi.org/10.3390/ani10060963.
22. Vollmer, W.; Blanot, D.; De Pedro, M.A. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews* **2008**, *32*, 149-167, https://doi.org/10.1111/j.1574-6976.2007.00094.x.
23. Sauvage, E.; Kerff, F.; Terrak, M.; Ayala, J.A.; Charlier, P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS microbiology reviews* **2008**, *32*, 234-258. https://doi.org/10.1111/j.1574-6976.2008.00105.x.
24. Fishovitz, J.; Hermoso, J.A.; Chang, M.; Mobashery, S. Penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus. *IUBMB Life* **2014**, *66*, 572-577, https://doi.org/10.1002/iub.1289. https://doi.org/10.1002/iub.1289.
25. Blázquez, B.; Llarrull, L.I.; Luque-Ortega, J.R.; Alfonso, C.; Boggess, B.; Mobashery, S. Regulation of the expression of the β-lactam antibiotic-resistance determinants in methicillin-resistant Staphylococcus aureus (MRSA). *Biochemistry* **2014**, *53*, 1548-1550. https://doi.org/10.1021/bi500074w.
26. Ryffel, C.; Kayser, F.H.; Berger-Bächi, B. Correlation between regulation of mecA transcription and expression of methicillin resistance in staphylococci. *Antimicrobial Agents and Chemotherapy* **1992**, *36*, 25-31. https://doi.org/10.1128/AAC.36.1.25.
27. Francois, P.; Bento, M.; Renzi, G.; Harbarth, S.; Pittet, D.; Schrenzel, J. Evaluation of three molecular assays for rapid identification of methicillin-resistant Staphylococcus aureus. *Journal of clinical microbiology* **2007**, *45*, 2011-2013. https://doi.org/10.1128/JCM.00232-07.
28. Wu, M.; Tong, X.; Liu, S.; Wang, D.; Wang, L.; Fan, H. Prevalence of methicillin-resistant Staphylococcus aureus in healthy Chinese population: a system review and meta-analysis. *PLoS One* **2019**, *14*, e0223599. https://doi.org/10.1371/journal.pone.0223599.
29. de Kraker, M.E.A.; Jarlier, V.; Monen, J.C.M.; Heuer, O.E.; van de Sande, N.; Grundmann, H. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. *Clinical Microbiology and Infection* **2013**, *19*, 860-868, https://doi.org/https://doi.org/10.1111/1469-0691.12028.
30. Garoy, E.Y.; Gebreab, Y.B.; Achila, O.O.; Tekeste, D.G.; Kesete, R.; Ghirmay, R.; Kiflay, R.; Tesfu, T. Methicillin-resistant Staphylococcus aureus (MRSA): prevalence and antimicrobial sensitivity pattern among patients—a multicenter study in Asmara, Eritrea. *Canadian Journal of Infectious Diseases and Medical Microbiology* **2019**. https://doi.org/10.1155/2019/8321834.

31. Guzmán-Blanco, M.; Mejía, C.; Isturiz, R.; Alvarez, C.; Bavestrello, L.; Gotuzzo, E.; Labarca, J.; Luna, C.M.; Rodríguez-Noriega, E.; Salles, M.J. Epidemiology of meticillin-resistant Staphylococcus aureus (MRSA) in Latin America. *International journal of antimicrobial agents* **2009**, *34*, 304-308. https://doi.org/10.1016/j.ijantimicag.2009.06.005.
32. Jiménez, J.N.; Ocampo, A.M.; Vanegas, J.M.; Rodriguez, E.A.; Mediavilla, J.R.; Chen, L.; Muskus, C.E.; A. Vélez, L.; Rojas, C.; Restrepo, A.V. CC8 MRSA strains harboring SCC mec type IVc are predominant in Colombian hospitals. *PLoS One* **2012**, *7*, e38576. https://doi.org/10.1371/journal.pone.0038576.
33. Falagas, M.E.; Karageorgopoulos, D.E.; Leptidis, J.; Korbila, I.P. MRSA in Africa: filling the global map of antimicrobial resistance. *PloS one* **2013**, *8*, e68024. https://doi.org/10.1371/journal.pone.0068024.
34. Bateman, A.; Rawlings, N.D. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends in biochemical sciences* **2003**, *28*, 234-237. https://doi.org/10.1016/S0968-0004(03)00061-6.
35. Oliveira, H.; Melo, L.D.; Santos, S.B.; Nóbrega, F.L.; Ferreira, E.C.; Cerca, N.; Azeredo, J.; Kluskens, L.D. Molecular aspects and comparative genomics of bacteriophage endolysins. *Journal of virology* **2013**, *87*, 4558-4570. https://doi.org/10.1128/JVI.03277-12.
36. Mayer, B.J.; Eck, M.J. SH3 domains: minding your p's and q's. *Current Biology* **1995**, *5*, 364-367. https://doi.org/10.1016/S0960-9822(95)00073-X.
37. Becker, S.C.; Foster-Frey, J.; Stodola, A.J.; Anacker, D.; Donovan, D.M. Differentially conserved staphylococcal SH3b_5 cell wall binding domains confer increased staphylolytic and streptolytic activity to a streptococcal prophage endolysin domain. *Gene* **2009**, *443*, 32-41. https://doi.org/10.1016/j.gene.2009.04.023.
38. Ponting, C.P.; Aravind, L.; Schultz, J.; Bork, P.; Koonin, E.V. Eukaryotic Signalling Domain Homologues in Archaea and Bacteria. Ancient Ancestry and Horizontal Gene Transfer. *Journal of Molecular Biology* **1999**, *289*, 729-745, https://doi.org/10.1006/jmbi.1999.2827.
39. Baba, T.; Schneewind, O. Target cell specificity of a bacteriocin molecule: a C-terminal signal directs lysostaphin to the cell wall of Staphylococcus aureus. *The EMBO journal* **1996**, *15*, 4789-4797. https://doi.org/10.1002/j.1460-2075.1996.tb00859.x.
40. Low, L.Y.; Yang, C.; Perego, M.; Osterman, A.; Liddington, R.C. Structure and lytic activity of a Bacillus anthracis prophage endolysin. *Journal of Biological Chemistry* **2005**, *280*, 35433-35439. https://doi.org/10.1074/jbc.M502723200.
41. Loessner, M.J.; Kramer, K.; Ebel, F.; Scherer, S. C-terminal domains of Listeria monocytogenes bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. *Molecular microbiology* **2002**, *44*, 335-349. https://doi.org/10.1046/j.1365-2958.2002.02889.x.
42. Fischetti, V.A. Bacteriophage lysins as effective antibacterials. *Current opinion in microbiology* **2008**, *11*, 393-400. https://doi.org/10.1016/j.mib.2008.09.012.
43. Díez-Martínez, R.; de Paz, H.; Bustamante, N.; García, E.; Menéndez, M.; García, P. Improving the lethal effect of Cpl-7, a pneumococcal phage lysozyme with broad bactericidal activity, by inverting the net charge of its cell wall-binding module. *Antimicrobial agents and chemotherapy* **2013**, *57*, 5355-5365. https://doi.org/10.1128/AAC.01372-13.
44. Fang, Y.; Fang, J. Discrimination of soluble and aggregation-prone proteins based on sequence information. *Molecular BioSystems* **2013**, *9*, 806-811. https://doi.org/10.1039/C3MB70033J.
45. Madhavan, A.; Sindhu, R.; Binod, P.; Sukumaran, R.K.; Pandey, A. Strategies for design of improved biocatalysts for industrial applications. *Bioresource technology* **2017**, *245*, 1304-1313. https://doi.org/10.1016/j.biortech.2017.05.031.
46. Chan, W.-C.; Liang, P.-H.; Shih, Y.-P.; Yang, U.-C.; Lin, W.-c.; Hsu, C.-N. Learning to predict expression efficacy of vectors in recombinant protein production. *BMC bioinformatics* **2010**, *11*, 1-12. https://doi.org/10.1186/1471-2105-11-S1-S21.
47. Musil, M.; Konegger, H.; Hon, J.; Bednar, D.; Damborsky, J. Computational design of stable and soluble biocatalysts. *Acs Catalysis* **2018**, *9*, 1033-1054. https://doi.org/10.1021/acscatal.8b03613.
48. Hebditch, M.; Carballo-Amador, M.A.; Charonis, S.; Curtis, R.; Warwicker, J. Protein–Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **2017**, *33*, 3098-3100. https://doi.org/10.1093/bioinformatics/btx345.

49. Hon, J.; Marusiak, M.; Martinek, T.; Kunka, A.; Zendulka, J.; Bednar, D.; Damborsky, J. SoluProt: prediction of soluble protein expression in Escherichia coli. *Bioinformatics* **2021**, *37*, 23-28. https://doi.org/10.1093/bioinformatics/btaa1102.
50. Raimondi, D.; Orlando, G.; Fariselli, P.; Moreau, Y. Insight into the protein solubility driving forces with neural attention. *PLoS computational biology* **2020**, *16*, e1007722. https://doi.org/10.1371/journal.pcbi.1007722.
51. Wang, X.; Li, L.; Li, S.; Huang, J.; Fan, Y.; Yao, Z.; Ye, X.; Chen, S. Phenotypic and molecular characteristics of Staphylococcus aureus and methicillin-resistant Staphylococcus aureus in slaughterhouse pig-related workers and control workers in Guangdong Province, China. *Epidemiology & Infection* **2017**, *145*, 1843-1851. https://doi.org/10.1017/S0950268817000085.
52. Ye, X.; Liu, W.; Fan, Y.; Wang, X.; Zhou, J.; Yao, Z.; Chen, S. Frequency-risk and duration-risk relations between occupational livestock contact and methicillin-resistant Staphylococcus aureus carriage among workers in Guangdong, China. *American Journal of Infection Control* **2015**, *43*, 676-681. https://doi.org/10.1016/j.ajic.2015.03.026.
53. Vanderhaeghen, W.; Piepers, S.; Leroy, F.; Van Coillie, E.; Haesebrouck, F.; De Vliegher, S. Invited review: effect, persistence, and virulence of coagulase-negative Staphylococcus species associated with ruminant udder health. *Journal of Dairy Science* **2014**, *97*, 5275-5293. https://doi.org/10.3168/jds.2013-7775.
54. Thorberg, B.-M.; Danielsson-Tham, M.-L.; Emanuelson, U.; Waller, K.P. Bovine subclinical mastitis caused by different types of coagulase-negative staphylococci. *Journal of dairy science* **2009**, *92*, 4962-4970. https://doi.org/10.3168/jds.2009-2184.
55. Supré, K.; Haesebrouck, F.; Zadoks, R.; Vaneechoutte, M.; Piepers, S.; De Vliegher, S. Some coagulase-negative Staphylococcus species affect udder health more than others. *Journal of dairy science* **2011**, *94*, 2329-2340. https://doi.org/10.3168/jds.2010-3741.
56. De Visscher, A.; Piepers, S.; Haesebrouck, F.; De Vliegher, S. Intramammary infection with coagulase-negative staphylococci at parturition: Species-specific prevalence, risk factors, and effect on udder health. *Journal of Dairy Science* **2016**, *99*, 6457-6469. https://doi.org/10.3168/jds.2015-10458.
57. Schleifer, K.H.; Kandler, O. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological reviews* **1972**, *36*, 407-477. https://doi.org/ 10.1128/br.36.4.407-477.1972.
58. Lu, J.Z.; Fujiwara, T.; Komatsuzawa, H.; Sugai, M.; Sakon, J. Cell wall-targeting domain of glycylglycine endopeptidase distinguishes among peptidoglycan cross-bridges. *Journal of Biological Chemistry* **2006**, *281*, 549-558. https://doi.org/10.1074/jbc.M509691200.
59. Gru¨ndling, A.; Schneewind, O. Cross-linked peptidoglycan mediates lysostaphin binding to the cell wall envelope of Staphylococcus aureus. *Journal of bacteriology* **2006**, *188*, 2463-2472. https://doi.org/10.1128/JB.188.7.2463-2472.2006.
60. Donovan, D.M.; Lardeo, M.; Foster-Frey, J. Lysis of staphylococcal mastitis pathogens by bacteriophage phi11 endolysin. *FEMS Microbiology Letters* **2006**, *265*, 133-139, https://doi.org/10.1111/j.1574-6968.2006.00483.x.
61. Donovan, D.M.; Dong, S.; Garrett, W.; Rousseau, G.M.; Moineau, S.; Pritchard, D.G. Peptidoglycan Hydrolase Fusions Maintain Their Parental Specificities. *Applied and Environmental Microbiology* **2006**, *72*, 2988-2996, https://doi.org/10.1128/AEM.72.4.2988-2996.2006.
62. De Jonge, B.; Chang, Y.-S.; Gage, D.; Tomasz, A. Peptidoglycan composition of a highly methicillin-resistant Staphylococcus aureus strain. The role of penicillin binding protein 2A. *Journal of Biological Chemistry* **1992**, *267*, 11248-11254. https://doi.org/10.1016/S0021-9258(19)49903-1.
63. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **2008**, *9*, 1-15. https://doi.org/10.1186/1471-2164-9-75.
64. 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. CDD: NCBI's conserved domain database. *Nucleic acids research* **2015**, *43*, D222-D226. https://doi.org/10.1093/nar/gku1221.
65. Wickham, H. *ggplot2: elegant graphics for data analysis*; springer: 2016.
66. Team, R.C. R: A language and environment for statistical computing. **2013**.

67. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **2013**, *30*, 772-780. https://doi.org/10.1093/molbev/mst010.

68. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **2010**, *5*, e9490. https://doi.org/10.1371/journal.pone.0009490.

69. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T.Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **2017**, *8*, 28-36. https://doi.org/10.1111/2041-210X.12628.

70. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold-Making protein folding accessible to all. **2021**. https://doi.org/10.1038/s41592-022-01488-1.

71. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Meng, E.C.; Couch, G.S.; Croll, T.I.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **2021**, *30*, 70-82. https://doi.org/10.1002/pro.3943.

72. Mónico, A.; Martínez-Senra, E.; Zorrilla, S.; Pérez-Sala, D. Drawbacks of dialysis procedures for removal of EDTA. *PloS one* **2017**, *12*, e0169843. https://doi.org/10.1371/journal.pone.0169843.

73. Ogle, D.; Ogle, M.D. Package 'FSA'. *CRAN Repos* **2017**, 1-206.