

Research Article

Collagen-Derived Cryptides: Machine-Learning Prediction and Molecular Dynamic Interaction Against *Klebsiella pneumoniae* Biofilm Synthesis Precursor

Ahmad Al-Khdhairawi¹, Siti Mariani Mhd-Marzuki¹, Zi-Shen Tan¹, Narin Shan¹, Danish Sanuri¹, Rahmad Akbar², Su Datt Lam³ and Fareed Sairi^{1*}

¹Department of Biological Science and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

²Department of Immunology, Oslo University Hospital Rikshospitalet and University of Oslo, Norway

³Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

*Corresponding author: fareed@ukm.edu.my

ABSTRACT

Collagen-derived cryptic peptides (cryptides) are biologically active peptides derived from the proteolytic digestion of collagen protein. These cryptides possess a multitude of activities, including antihypertensive, antiproliferative, and antibacterial. The latter, however, has not been extensively studied. The cryptides are mainly obtained from the protein hydrolysate, followed by characterizations to elucidate the function, limiting the number of cryptides investigated within a short period. The recent threat of antimicrobial resistance microorganisms (AMR) to global health requires the rapid development of new therapeutic drugs. The current study aims to predict antimicrobial peptides (AMP) from collagen-derived cryptides, followed by elucidating their potential to inhibit biofilm-related precursors in *Klebsiella pneumoniae* using in silico approach. Therefore, cryptides derived from collagen amino acid sequences of various types and species were subjected to online machine-learning platforms (i.e., CAMPr3, DBAASP, dPABBs, Hemopred, and ToxinPred). The peptide-protein interaction was elucidated using molecular docking, molecular dynamics, and MM-PBSA analysis against MrkH, a *K. pneumoniae*'s transcriptional regulator of type 3 fimbriae that promote biofilm formation. As a result, six potential antibiofilm inhibitory cryptides were screened and docked against MrkH. All six peptides bind stronger than the MrkH ligand (c-di-GMP; C2E).

Key words: Collagen, Antibiofilm peptide (AMP), *Klebsiella pneumoniae*, molecular docking, MrkH, Type 3 Fimbriae

Article History

Accepted: 28 November 2022

First version online: 26 December 2022

Cite This Article:

Al-Khdhairawi, A., Mhd-Marzuki, S.M., Tan, Z.S., Shan, N., Sanuri, D., Akbar, R., Lam, S.D. & Sairi, F. 2022. Collagen-derived cryptides: machine-learning prediction and molecular dynamic interaction against *Klebsiella pneumoniae* biofilm synthesis precursor. *Malaysian Applied Biology*, 51(5): 59-75. <https://doi.org/10.55230/mabjournal.v51i5.2351>

Copyright

© 2022 Malaysian Society of Applied Biology

INTRODUCTION

Antimicrobial resistance (AMR) is widely recognized as a serious threat to global health. By 2050, AMR could cause the deaths of 10 million people (O'Neill, 2016). and increase the burden of drug-resistant diseases caused by various pathogens in several countries (Murray *et al.*, 2022). The presence of AMR has compromised the effectiveness of existing treatments and is exacerbated by the misuse of antibiotics and the lack of new antibiotic developments (Mobarki *et al.*, 2019). Therefore, tackling AMR requires new tools or compounds to accelerate research and development of new antimicrobial therapeutic classes and raise awareness of antibiotics in healthcare and agriculture. One of the compounds being explored is antimicrobial peptides (AMPs). AMPs are parts of the innate immune system with antimicrobial properties to kill pathogenic microorganisms. A derivative of AMP, known as cryptic antimicrobial peptide (antimicrobial cryptide), is a unique class of AMP derived from proteolytically hydrolyzed protein.

The cryptides in collagen hydrolysate have been shown to have angiogenic, mitogenic (Banerjee & Shanthy, 2016), and antibacterial properties (Ennaas *et al.*, 2016; Atef *et al.*, 2021). The administration of collagen peptides significantly

improved exercise-athletic injury pain (Dressler et al., 2018; Zdzieblik et al., 2021) and accelerated healing in patients with Achilles tendonitis (Praet et al., 2019). Baehaki et al., (2016) reported that fish skin cryptides produced with bacterial collagenase inhibited angiotensin I-converting enzymes and were anti-cancer. On the other hand, cryptides derived from collagen showed antimicrobial activity against *Salmonella* spp., *Escherichia coli*, and *Staphylococcus aureus* (Atef et al., 2021). However, the antimicrobial potential of collagen-derived cryptides still needs to be extensively researched despite the protein diversity from different organisms.

More than hundreds of collagen proteins curated in the UniProtKB database are potential reservoirs for numerous biologically active cryptides. To date, cryptides have been identified and purified using mass spectrometry to elucidate their function (Baehaki et al., 2016; Ennaas et al., 2016; Atef et al., 2021). However, this approach is expensive, time-consuming, and limited to the characterization of a small number of peptides. Therefore, rapid screening and prediction strategies using machine learning (ML) are the best options to narrow down AMP before downstream analysis.

Several online prediction tools based on ML have been developed to speed up the prediction of AMP, e.g., CAMPr3 (Waghu et al., 2016), AMPPred (Pan et al., 2012), AntiBP (Lata et al., 2007), DRAMP (Shi et al., 2022). The quality of prediction of AMP is further improved by other ML tools that predict antibiofilm (dPABBs) (Sharma et al., 2016), hemolytic (Hemopred, DBAASP) (Pirtskhalava et al., 2021; Win et al., 2017), cell-penetrating (CPPPred) (Holton et al., 2013), potential target microorganisms (DBAASP) and toxicity (ToxinPred) (Gupta et al., 2013). In light of these findings, the current study focuses on predicting antimicrobial cryptides from collagen proteins and potential peptide-protein interactions with MrkH, a transcriptional activator for *K. pneumoniae* type III fimbriae.

The type III fimbriae, encoded by the mrk gene cluster, are an important precursor for the bacteria to initiate biofilm synthesis (Struve et al., 2009). In addition to the capsule, biofilm-producing *K. pneumoniae* are known to be more resistant to antibiotics than strains that cannot form a biofilm (Zheng et al., 2018). Uniquely in *K. pneumoniae*, the expression of the gene cluster is regulated by the MrkH protein (Wilksch et al., 2011). The protein also autoregulates its expression when activated by a global regulator, cyclic-di-GMP (C2E) molecules (Schumacher & Zeng, 2016). Inhibition of the MrkH protein is thought to disrupt bacterial attachment and thus reduce biofilm synthesis (Roy et al., 2018). Therefore, we aim to screen collagen-derived cryptides with antimicrobial

and antibiofilm potential against *K. pneumoniae* and determine the candidates' potential as MrkH inhibitors by molecular docking and molecular dynamic analyses.

MATERIALS AND METHODS

Datasets

A total of 21 collagen protein sequences from various organisms were downloaded from UniProtKB. As for positive and negative peptide references, 2754 antimicrobial peptide (AMP) sequences were curated from the Antimicrobial Peptide Database (APD) and 5850 non-AMP peptide sequences were obtained from UniProtKB, respectively. Non-AMPs are peptides that are recorded without antimicrobial properties such as "antimicrobial", "antibacterial", "antifungal" and "antiviral". The crystal structure of MrkH protein (PDB ID: 5KEC, 1.95 Å) was retrieved from the Protein Data Bank (PDB) (Berman et al., 2000), while cyclic-di-GMP (C2E) from native protein structure (5KGO, 2.90 Å).

Cryptides generation and physicochemical elucidation

A total of 34116 collagen-derived cryptic peptides (cryptides) were generated in-silico by proteolytically digesting all 21 collagen proteins with enzymes and chemicals available in Peptide Cutter (https://web.expasy.org/peptide_cutter/) (Gasteiger et al., 2005). The physicochemical properties of all peptides (AMP, Non-AMP, and collagen-derived cryptides) were elucidated based on amino acid composition using R Studio's Peptide package (Osorio et al., 2015). The physicochemical properties are net charge, Grand average of hydropathicity index (GRAVY), and Boman index.

Antimicrobial peptide (AMP) prediction and scoring

All peptides (AMP, Non-AMP, & collagen-derived cryptides) were analyzed by the CAMPr3 AMP prediction tool (<http://www.camp3.bicnirrh.res.in/prediction.php>) using Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN) and Discriminant Analysis (DA) features (Waghu et al., 2016). Subsequently, the collagen-derived cryptic peptides were scored according to the number of features that predict the peptide as AMP (Score value: 0-4). The cryptides were compared with AMP and Non-AMP using box-plot analysis and their physicochemical properties. The peptide sequences are also uploaded into the CAMP Sign database to cross-reference the sequences with known AMP families that have been experimentally verified to identify the possibility of known AMPs originating from the

collagen protein.

Collagen-derived cryptides biological activity prediction

The biological activity of collagen-derived cryptides against *K. pneumoniae* was predicted using DBAASP v3.0 (<https://dbaasp.org/home>) (Pirtskhalava et al., 2021), while antibiofilm peptide prediction was carried out by dPABBs (<https://ab-openlab.csir.res.in/abp/antibiofilm/>). Hemolytic activity of collagen-derived cryptides was predicted using DBAASP v3.0 and HemoPred (<http://codes.bio/hemopred/>) (Pirtskhalava et al., 2021), while toxicity was predicted by ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred/>) (Gupta et al., 2013). Selected collagen-derived cryptides were then subjected to structure prediction and docking analysis against MrkH protein.

Statistical analysis

A non-parametric statistical analysis using Kolmogorov-Smirnov (KS) test was conducted to determine the significant differences between AMP and NAMP. On the other hand, a non-parametric one-way analysis of variance (ANOVA) was performed to compare the significance of the cryptides group in AMP scores with AMP followed by the Kruskal-Wallis test for multiple comparison analysis. The significant difference between data was presented as a *p-value* (<0.05).

Peptide structure prediction and validation

PEP-FOLD3 was used to predict the three-dimensional structure of a collection of peptides (<http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/>) that take a de novo approach (Lamiable et al., 2016). Predicted peptide profiles were assembled using a coarse-grained force field and the structural alphabet (SA). SOPEP energy and TM score were used to sort the clusters, and the Ramachandran plot was used to verify the structure from the perspective of phi and pi angles.

Receptor and peptides preparation for molecular docking

YASARA Structure software (Land & Humble, 2018) was used to optimize the MrkH receptor protein. The default energy minimization parameter was used to minimize the energy of all 3D structures of molecules and selected peptide molecules before docking. The protein chain was edited for missing hydrogen atoms, bond orders, and optimization of hydrogen bonds. The preparation process of the protein continued until it attained a minimized state with the help of the NOVA force field (Krieger et al., 2002).

Molecular docking

HADDOCK2.4 webserver (Honorato et al., 2021; van Zundert et al., 2016) was used for molecular docking of the candidate, considering the confined co-crystalline binding site as a chemical search space. The HADDOCK method created the score function for peptides docked inside the active site that was used in this study for molecular docking. The binding pocket was generated using C2E ligand active site residues 67, 69, 73, 74, 107, 108, 109, 110, 113, 140, 141, 142, 146, 147, 184, 185, 187, 203, 205, 206, 207 to select interaction area of monomers. HADDOCK configuration parameters are set to defaults. The screened compounds' PRODIGY binding energy (kcal/mol) (Vangone et al., 2019) was used as a rank score.

Molecular dynamics

Molecular dynamics simulations (MD) were performed using the YASARA Structure (version 21.12.19) protocol (Krieger & Vriend, 2015) to gain a better understanding of the stability of protein-ligand complexes. All simulations were carried out on a custom-built workstation running Linux Ubuntu 20.04.5 (Focal Fossa) with AMD Ryzen Threadripper 3960X 24-core processor (4.5 GHz), 64 Gb of random access memory (RAM), and an NVIDIA GeForce RTX 3080 graphics processing unit (GPU) (10 Gb). The AMBER14 force field (Maier et al., 2015) was used for the MD simulations. Initial energy minimization was performed using the steepest descent algorithm. The MD simulations were performed for amino acid residues with the standard physiological pH (7.4). Water molecules were successfully introduced into the system at constant temperature and pressure conditions. A counter ion (Na⁺ or Cl⁻) with a concentration of 0.9% was added to maintain the neutral state of the systems. The Berendsen barostat technique (Berendsen et al., 1984) was used to maintain the pressure value at 1 atm. The long-range Coulomb forces were calculated using the particle-mesh Ewald (PME) method (Krieger et al., 2006; Krieger & Vriend, 2015). The cut-off radius was set to 8 Å for the non-bonded interactions. The temperature was kept at 298 K using the Langevin thermostat method (Izaguirre et al., 2001). The periodic boundary conditions were also considered. In each case, the cubic simulation cell was chosen 20 Å larger than the protein-ligand complexes studied. A normal simulation speed was maintained with a time step of 1.25 fs. At an integration step of 2 fs, all bonds, including hydrogen bonds, were constrained using the algorithm SHAKE (Elber et al., 2011). In the final step, the production steps were run over a simulation time of 100 ns. Snapshots of the simulation trajectory were taken every 100 ps. The

simulation steps were executed using a preloaded macro (md_run.mcr) from the YASARA package. Root mean square deviation (RMSD), the radius of gyration (Rg), root-mean-square fluctuation (RMSF), and solvent accessible surface area (SASA) were used to analyze the best compounds at the end of the 100 ns simulations MD.

MM/PBSA binding energy calculations

The studied peptides' binding free energies against the MrkH protein were calculated using the mechanics Poisson-Boltzmann surface approach (MM/PBSA) by YASARA built-in macro md_analyzebindenergy.mcr. The theory of nuclear physics is used as the basis for the approach (Blatt & Weisskopf, 1979). Similar methods to MM/PBSA have been reported by (Swanson *et al.*, 2004). By subtracting the energy of the system, the binding energy can be approximated from the energy at an infinite distance. The interaction is more favorable the higher the binding energy. Binding free energy was calculated with the solvation of peptide, complex, and free protein for the MrkH form complexes using AMBER 14 as a force field, where more positive energies indicate better binding. The binding free energy (kcal/mol) for the MM/PBSA was calculated according the following equation:

$$\Delta G \text{ binding} = \text{EpotRecept}(i) + \text{EsolvRecept}(i) + \text{EpotLigan} + \text{EsolvLigand} - [\text{EpotComplex}(i) + \text{EsolvComplex}(i)].$$

where *i* is the position number, "Epot" potential energy, and "Esolv" is the solvation energy, free protein (EpotRecept) and (EsolvRecept), free ligand (EpotLigan) and (EsolvLigand), while the potential energy for the complex is (EpotComplex) and the solvation energy for the complex (EsolvComplex).

RESULTS

Collagen-derived cryptides generation

A total of 21 collagen protein sequences from *Homo sapiens* (humans), *Mus musculus* (mice), and *Gallus gallus* (chicken) were selected for cryptides generation using the Peptide Cutter platform. The collagen proteins include fibrillar collagens, fibril-associated collagens with interrupted triple helices (FACIT), fibril-associated collagen, multiplexing, and network-forming collagen. In this study, a total of 34 116 cryptides were generated from the proteins. Out of 21 proteins, three collagens from *Gallus gallus* (CO6A3, COCA1, COEA1) generated 5875 cryptides, nine collagens from *Homo sapiens* (CO5A1, CO6A3, CO6A5, CO6A6, CO7A1, COCA1, COEA1, COIA1, COOA1) generated 14 026 cryptides and another nine from

Mus musculus (CO5A1, CO6A4, CO6A5, CO6A6, CO7A1, COCA1, COEA1, COIA1, COOA1) generated 14 215 cryptides.

Neutrophil elastase, thermolysin, and low-specificity chymotrypsin generated 3256, 3163, and 2849 cryptides. Neutrophil elastase is an enzyme produced by neutrophils that play a vital role in the body's defense system, and chymotrypsin is a digestive enzyme produced by the digestive system. The former degrade microorganisms or any foreign substances engulfed by neutrophils and act in concert with reactive oxygen species (Korkmaz *et al.*, 2010). On the other hand, thermolysin is an enzyme secreted by the *Bacillus thermoproteolyticus* to degrade proteins and peptides as nutrients for bacterial growth (Adekoya & Sylte, 2009).

AMP prediction and scoring

Prediction of collagen-derived cryptides by CAMPr3 using SVM, RF, ANN, and DA resulted in more than 14000 cryptides (41%) predicted as AMP by at least one classifier. In Table 1, predicted cryptides with antimicrobial properties were distributed according to the classifier using AMP scoring values of 0-4. Overall, the number of cryptides decreased as the AMP scoring value increased, suggesting that cryptides with a score of three and four are highly potential AMP. Out of 41% of potential AMP, only 4.35% of the cryptides scored three and four, represented by 950 and 535 cryptides. Among the collagen types, Collagen Type VI (Co6) contributes the most cryptides with an AMP score of three and four between the three species. Co6 is the collagen that forms the tissue and connective tissue, often associated with the basal membrane. The potential antimicrobial activity of Co6 was previously demonstrated against Groups A, C, and G Streptococci (Abdillahi *et al.*, 2012), *Moraxella catarrhalis* (Abdillahi *et al.*, 2015), *Pseudomonas aeruginosa*, *E. coli*, and *S. aureus* (Abdillahi *et al.*, 2018).

Among Co6-derived cryptides, nine were similar to three AMP families curated in the CAMP Sign database (Table 2). The AMP families are Uperin, Bacteriocin, Latarcin, and Ascaphin. Uperin, which is active against Gram-negative bacteria, was isolated from frogs' dorsal gland or skin from the family Uperoleia (Bradford *et al.*, 1996), while Ascaphin from the Ascaphus family (Conlon *et al.*, 2004). On the other hand, Bacteriocin is a family of AMPs produced by probiotic bacteria and capable of inhibiting the growth of pathogens and other strains (Dobson *et al.*, 2012). Another four cryptides derived from human Co6 are grouped into the Latarcin family. Latarcin is a peptide isolated from the venom of a group of spiders that exhibit antimicrobial activity by penetrating the microbial cells (Budagavi &

Table 1. AMP scoring distribution of collagen-derived cryptides according to species and collagen types.

Species	Type	Number of cryptides per amp scoring					Total
		0	1	2	3	4	
<i>Gallus gallus</i>	co6a3	1259	676	168	86	43	2232
	coca1	1515	554	167	63	29	2328
	coea1	813	355	101	27	19	1315
		3587	1585	436	176	91	5875
<i>Homo sapiens</i>	co5a1	573	449	97	32	10	1161
	co6a3	1265	699	182	76	64	2286
	co6a5	1152	528	180	61	50	1971
	co6a6	1001	497	178	51	31	1758
	co7a1	1029	705	118	26	17	1895
	coca1	977	335	100	38	27	1477
	coea1	836	308	76	26	11	1257
	coia1	597	304	109	32	13	1055
	cooa1	611	378	106	42	29	1166
		8041	4203	1146	384	252	14026
<i>Mus musculus</i>	co5a1	597	426	100	31	11	1165
	co6a4	982	450	120	69	34	1655
	co6a5	1200	537	165	55	39	1996
	co6a6	951	476	182	66	29	1704
	co7a1	1036	701	121	30	14	1902
	coca1	1529	536	181	46	32	2324
	coea1	809	321	90	32	8	1260
	coia1	662	319	74	22	11	1088
	cooa1	603	380	85	39	14	1121
		8369	4146	1118	390	192	14215
Total		19997	9934	2700	950	535	34116

Table 2. Cryptides similarity within CAMP sign AMP family

No.	Cryptide	Family
1	co6a3_chick_19	Uperin
2	co6a3_human_940	Bacteriocin
3	co6a3_human_1545	Bacteriocin
4	co6a5_human_1375	Latarcin
5	co6a5_human_1596	Latarcin
6	co6a5_human_1884	Latarcin
7	co6a5_human_1970	Latarcin
8	co6a5_mouse_98	Ascaphin
9	co6a5_mouse_1529	Ascaphin

Chugh, 2018). Meanwhile, most peptides with AMP scores of three and four had no similarity to AMP from the database, indicating that the probability of finding a novel AMP candidate from this study is high.

Collagen-derived antimicrobial cryptides characterization

Prediction using separate machine-learning classifiers provides a limited consensus of the peptide potential as AMP. Thus, a scoring function was employed by summation of the classifier resulting in AMP to improve the prediction consensus for each cryptides. The net charge,

GRAVY, and Boman Index distribution of collagen-derived cryptides were compared to AMP and non-AMP (NAMP) from databases (Figure 1). As a reference, the differences in net charge, GRAVY, and Boman index between AMP and NAMP are significantly different ($p < 0.0001$) (Table 3). As for the Kolmogorov-Smirnov test, the D values for the net charge, GRAVY, and Boman Index were 0.4366, 0.2777, and 0.2887, respectively. D value represents the maximum distance between the two samples' cumulative distribution function (CDF). As the D value is close to 0, the similarity of the data distribution between the two samples increases. Thus, following previous studies, net

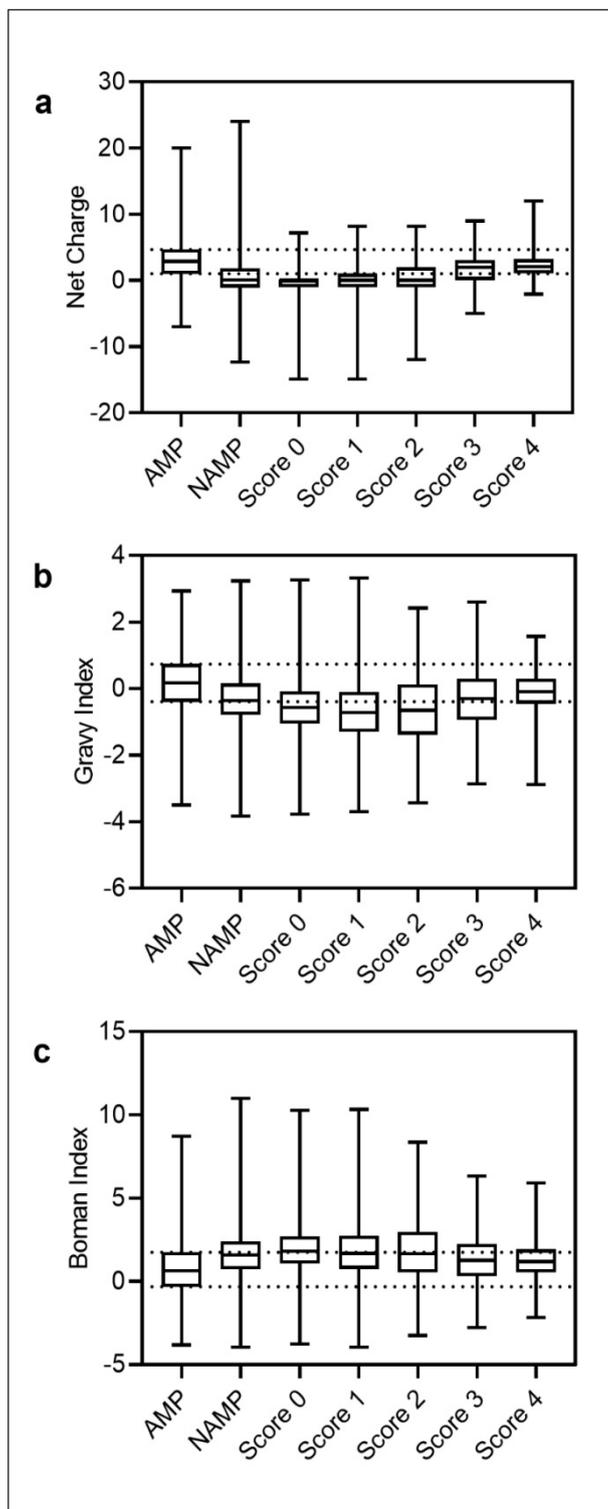


Fig. 1. Box-plot distribution for AMP, NAMP, and collagen-derived cryptides with scores for a) Net charge, b) GRAVY and c) Boman Index.

Table 3. Comparison of AMP and non-AMP parameters

Parameter	Net charge		GRAVY		Boman Index	
	AMP	Non-AMP	AMP	Non-AMP	AMP	Non-AMP
Minimum	-7	-12	-3.5	-3.8	-3.8	-3.9
The first quartile (Q1)	1	-1	-0.4	-0.8	-0.3	0.8
Median	3	0	0.2	-0.4	0.6	1.6
Third Quartile (Q3)	5	2	0.7	0.2	1.8	2.4
Maximum	20	24	2.9	3.2	8.7	11.0
<i>P-value</i>	< 0.0001		< 0.0001		< 0.0001	
Kolmogorov-Smirnov D value	0.4366		0.2777		0.2887	

charge demonstrated the most crucial features of AMP compared to GRAVY and Boman index (Torrent *et al.*, 2011; Chung *et al.*, 2020). Therefore, cryptides with all three values within 50% of AMP (quartile 1 and quartile 3 value) are considered potential AMP.

In Figure 1a, 50% of AMP within the first and third quartile has a net charge value between 1 and 5, respectively. The median for AMP and NAMP was 3 and 0, respectively. As depicted in Figure 1a, most cryptides with a score of three and four fall within 50% of AMP net charge, where both groups also have a median closest to the AMP. However, a one-way ANOVA test reveals that only the latter are not significantly different (P value > 0.9999), suggesting that cryptides scored four as highly potential AMP. Many AMPs discovered are cationic, with a net charge in the range of 2 to 9 (Porto *et al.*, 2012), while less than 5% of AMPs have an anionic net charge (Chung *et al.*, 2020). Net charge plays a crucial role in determining the antimicrobial properties of AMP and the differences in membrane structure between prokaryotic cells and eukaryotic cells make the cationic properties of an AMP an advantage in terms of antimicrobial efficacy (Yeaman & Yount, 2003). Eukaryotic cell membranes are neutral, while prokaryotic cell membranes are negatively charged. This allows the cationic AMP to bind to the negatively charged bacterial cell wall through electrostatic interactions without affecting the host cell. However, previous studies have also reported that the higher the net charge value of AMP the higher the probability of hemolytic activity of the peptide on human red blood cells. An increase in net charge is likely to cause a tendency to produce pores on eukaryotic cell membranes (Jiang *et al.*, 2008)

Unlike net charge, the GRAVY and Boman Index were inconclusive as depicted in Figure 1b. Although AMP scoring of cryptides that scored four had the median closest to AMP for both GRAVY and Boman Index, the p -value is <0.0003 and <0.0001, respectively. GRAVY is one of the main characteristics of AMP interaction with the microbial membrane (Mao *et al.*, 2013). In the current study,

50% of AMP GRAVY value is between -0.4 and 0.7, while the cryptides scored four had -0.4 and 0.3. The value is slightly below the optimal range (0 to 16.6) suggested by Chen *et al.*, (2007). Optimal GRAVY values improve the AMP penetration of the prokaryotic cell membranes (Bahar & Ren, 2013). On the other hand, the Boman Index value for the AMP median was lower (0.6) compared to NAMP (1.6), which contradicts the previous report (Qutb *et al.*, 2020). In Figure 1c, 50% of AMP recorded a value between -0.3 and 1.8, while NAMP was between 0.8 and 2.4. The collagen-derived cryptides distribution was between 0.5 and 1.9, which significantly differed when compared to AMP and NAMP (p <0.0001). The Boman index measures the ability of peptides to perform biological interactions through the affinity of peptides to proteins. Higher Boman index values are preferable for AMPs because they have higher binding potential to cell membrane proteins (He *et al.*, 2018). Nevertheless, all three parameters are used to narrow down potential antimicrobial cryptides by referencing the first and third-quartile values of net charge, GRAVY, and Boman Index. Based on the box-plot results, 256 collagen-derived cryptides were narrowed down according to AMP scoring 3 and 4, followed by having net charge (1 to 5), GRAVY (-0.4 to 0.7), and Boman Index (-0.3 to 1.8) value.

All 256 collagen-derived cryptides were then predicted for hemolytic activity (HemoPred and DBAASP v3), antibiofilm property (dPABBs), toxicity (ToxinPred), and potential inhibitory activity against *K. pneumoniae* (DBAASP v3). Out of 256 candidates, 24 collagen-derived cryptides were selected for their predicted activity against *K. pneumoniae* and antibiofilm activity without hemolytic and toxin properties (Table 4).

3-dimensional (3-D) structure prediction and optimization of collagen-derived cryptides

The 3-D structures of all 24 collagen-derived cryptides were predicted with PEPFOLD3. The cryptides are mostly α -helix (18 cryptides), β -sheet

Table 4. Physicochemical properties of collagen-derived cryptides with potential antimicrobial and antibiofilm activity against *K. pneumoniae*

ID	Source	Structure	Sequence	Peptide length	Net charge	Boman Index	GRAVY
Peptide 1	Human Co5a1	β -sheet, turn	GKWHRIALSVHKKNVTLILD	20	3	1.2	-0.1
Peptide 2	Human Co5a1	β -sheet, turn	DGKWHRIALSVHKKNVTLIL	20	3	1.2	-0.1
Peptide 3	Mouse co5a1	β -sheet, turn	GKWHRIALSVYKKNVTLILD	20	3	1.0	0.0
Peptide 4	Chicken co6a3	α -helix, turn	EVAQKGVKVFAVGVRNI	17	2	0.7	0.5
Peptide 5	Human co6a3	α -helix, turn	DVSLALTQRGVKFAVGVRNI	21	2	0.9	0.7
Peptide 6	Human co6a3	α -helix, turn	EIRYGVVALKQASVFSFGLGAQAASRA	27	2	0.6	0.5
Peptide 8	Human co6a3	α -helix	IRYGVVALKQASVFSFGLGAQAASRAE	27	2	0.6	0.5
Peptide 9	Mouse co6a4	α -helix, turn	EKGSRPHRGVQQIAVVII	18	2	1.7	-0.1
Peptide 10	Mouse co6a4	α -helix, turn	ENVLLTAVLPRRSRVLYAIVAS	22	2	1.0	0.7
Peptide 11	Mouse co6a5	α -helix, turn	EFVKTVLRAKCCQGYVVFVISLGSTQR	27	3	0.9	0.5
Peptide 12	Human co6a6	α -helix	EARGSRNLKGVQVLVIT	19	2	1.3	0.2
Peptide 13	Human co6a6	α -helix, turn	KGVKGAKGLASFSTCELIQYVR	22	3	0.9	0.1
Peptide 14	Mouse co6a6	turn (extend)	GSKVPCHLVLTNGMSR	17	2	0.8	0.3
Peptide 15	Mouse co6a6	α -helix, turn	EARGSRNLKGVQVLVIT	19	2	1.2	0.2
Peptide 16	Human co7a1	α -helix, turn	DTAAQRLKGQGVKLFVAVGIKNA	22	3	1.1	-0.1
Peptide 17	Human co7a1	α -helix, turn	TAAQRLKGQGVKLFVAVGIKNAD	22	3	1.1	-0.1
Peptide 18	Mouse co7a1	β -sheet, turn	DRVFLPRLTRPGVPKVCILIT	21	3	1.1	0.5
Peptide 19	Mouse co7a1	β -sheet, turn	RVFLPRLTRPGVPKVCILITD	21	3	1.1	0.5
Peptide 20	Chicken coca1	α -helix, turn	DAKELKLIASQPSLKHVFNVANF	23	1	1.0	0.0
Peptide 21	Chicken coca1	α -helix, turn	ELKLIASQPSLKHVFNVANF	20	1	0.5	0.3
Peptide 22	Human coca1	α -helix, turn	DGYEILGKLLKGERKSAAFQIQSF	24	1	1.5	-0.3
Peptide 23	Human coca1	α -helix, turn	GYEILGKLLKGERKSAAFQIQSFD	24	1	1.5	-0.3
Peptide 24	Chicken coea1	α -helix, turn	EAGMRKGIPKVLVIT	16	2	0.2	0.6
Peptide 25	Chicken coea1	α -helix, turn	RKGIPKVL	8	3	1.4	-0.2

(5 cryptides), and one extended structure (Table 4). Furthermore, all 24 cryptides 3D structures were validated by having more than 90% Ramachandran score. The secondary structure of an antimicrobial peptide is closely related to its function, e.g., binding to bacterial cell wall membranes (Mahlpuu *et al.*, 2016) and binding to intracellular receptors (Hourri & Mechler, 2020).

Molecular docking analysis

The peptide-MrkH interaction of all 24 collagen-derived cryptides was compared with the MrkH ligand (c-di-GMP) using the HADDOCK2.4 web

server. The results of all 24 candidates were promising, as the binding interaction with the MrkH cavity was stronger than C2E. As a reference, the binding affinity of C2E to the MrkH cavity is -6.9 kcal/mol, while the cryptides have a stronger affinity (Table 5). As summarised in Table 5, Peptides 6, 8, 15, 18, 19, and 24 achieved a binding affinity in the range of -9.3 to -11.3 (kcal/mol).

The peptide with the highest docking score was Peptide 18 with a binding energy of -11.3 kcal/mol, forming several hydrogen bonds with HIS120 and one hydrogen bond with the following residues: ARG117, ASN35, and GLN203. The

Table 5. Peptides with binding affinity score predicted by PRODIGY

Structure	ΔG (kcal mol ⁻¹)
C2E	-6.9
Peptide 6	-10.9
Peptide 8	-11
Peptide 15	-10.7
Peptide 18	-11.3
Peptide 19	-10.4 </td
Peptide 24	-9.3

For more clarity, 2D representations of the formed bonds in the six protein-peptides complexes are illustrated in Figure 2. All six peptides were placed deep in the active site of the MrkH protein as shown in Figure 3.

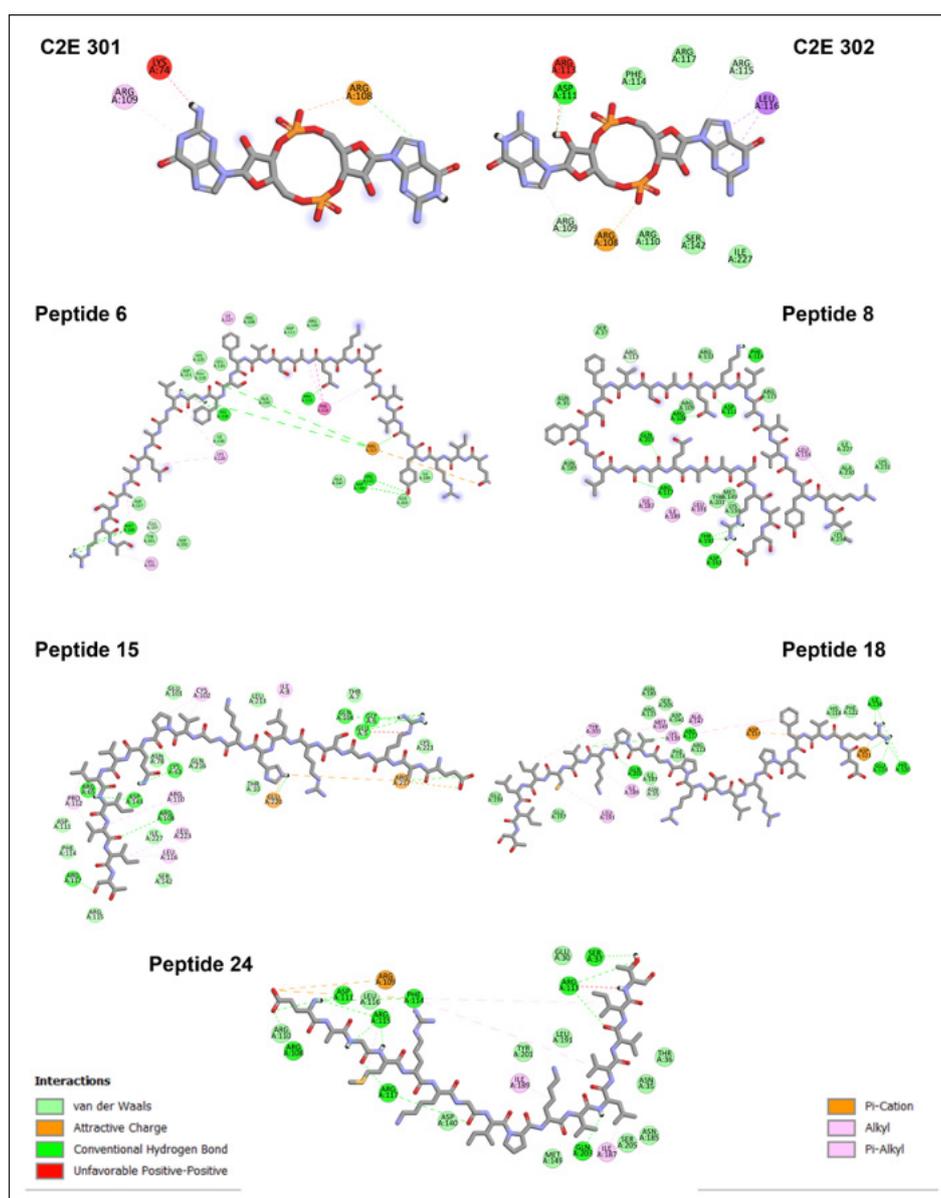


Fig. 2. Two-dimensional binding interactions of selected peptides and native ligand (generated by Discovery studio 2D visualize) with active site residues of 5KEC.

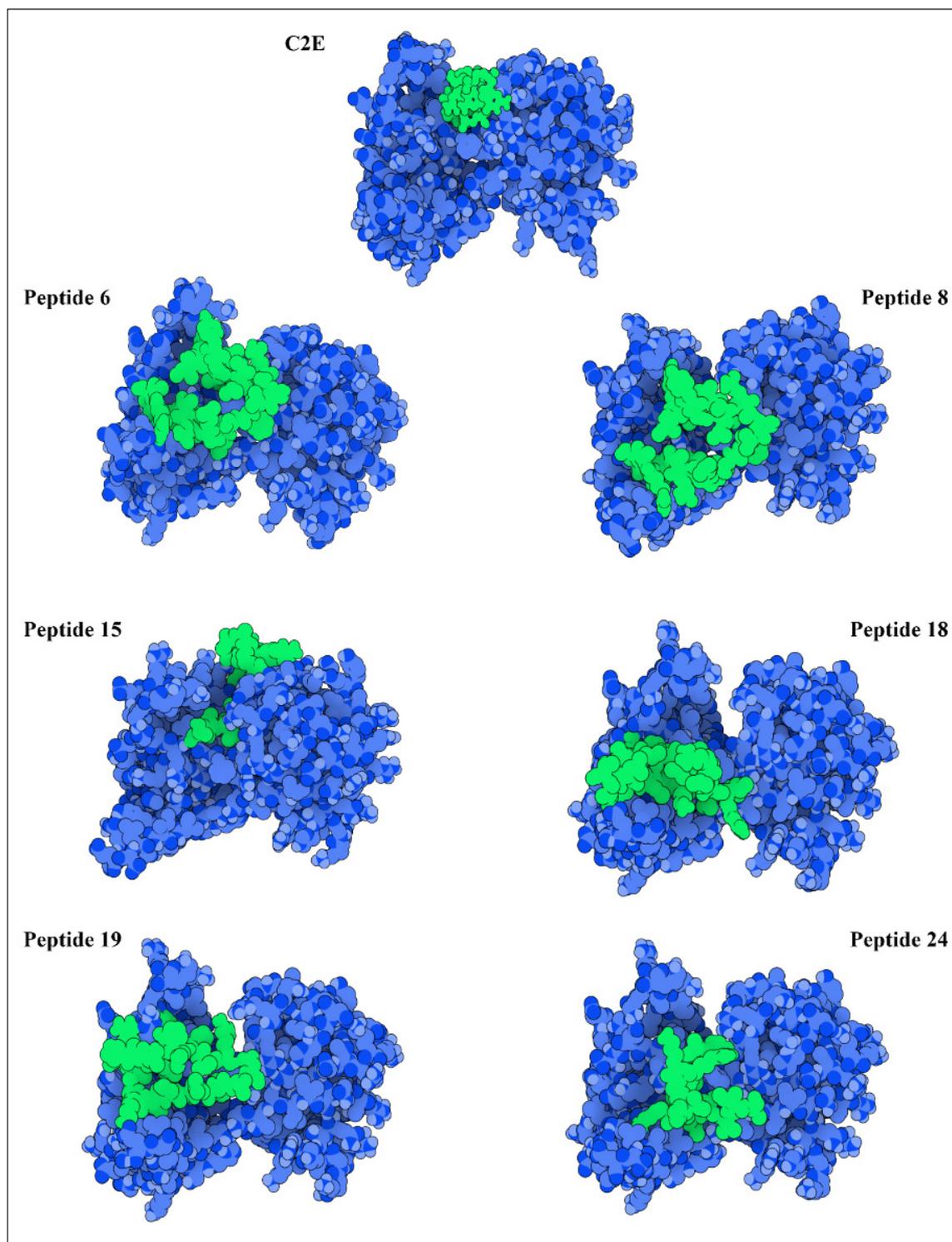


Fig. 3. The binding poses of selected peptides in the active site cavity of MrkH.

second top-ranked peptide is Peptide 8 with a binding energy of -11.0 kcal/mol, forming several hydrogen bonds with ARG108, ARG117, THR190 and one hydrogen bond with ARG109, ASP192, PHE114, and GLN203. Peptide 6 formed slightly more hydrogen bonds compared to Peptide 18 and Peptide 8, but with different interacting residues, namely ARG117, ARG113, ARG115, HIS118, ASP140, and GLN203. In the case of Peptide 15, thirteen hydrogen bonds were observed, one for THR10, ARG65, ASN78, ARG108, ARG117, ARG217, GLU220, and two in GLY6, LYS63, and ASP143. With a difference of 0.3 kcal/mol from Peptide 15, Peptide 19 formed eight hydrogen bonds with ARG113, ARG117, ARG109, GLU201, TYR201, and LEU226. Peptide 24 achieved a docking score of -8.0 kcal/mol with fourteen hydrogen bonds with SER37, ARG113, ARG115, ARG117, ASN35, ARG108, ASP111, PHE114, ILE189, and GLN203. In addition, different types of interactions were observed, such as pi-ion, pi-alkyl, pi-sigma, and pi-pi stacked. Docking of the native ligand C2E onto the prepared and optimized MrkH receptor model produced results consistent with the known interactions contributing to the stability of the complex in vitro (Schumacher & Zeng, 2016). In particular, the docked ligand was able to simulate interactions with the following residues: ARG108, ARG109, and ARG113.

Different types of interactions between peptide atoms and protein residues identify the extent of inhibition of the target protein (Martins *et al.*, 2021). Hydrogen bonding (HB) is considered an important of these interactions as it provides noticeable protein-ligand stability (Hubbard & Kamran Haider, 2010). Moreover, other types of interactions, such as pi-alkyl, pi-cation, and pi-sigma, were also observed.

Molecular dynamics (MD) simulations

The purpose of applying molecular dynamics (MD) simulations is attributed to its main role in investigating conformational stability and obtaining reliable results about the studied peptides' behavior inside the active site of MrkH protein. The best peptides were selected according to the obtained results from the molecular docking study and inspection of molecular properties. Six peptide-receptor complexes were studied using MM/PBSA binding energy calculations over 100 ns MD simulations to predict the peptide's binding behavior to the receptor. The calculated MM/PBSA binding energies of the six simulation complexes 100 ns were summarised in Table 6.

Although MM/PBSA still demonstrated that the peptide's binding affinity is better than the control ligand, the peptide ranking is different compared to the docking analysis. Previously, the docking results suggested that Peptide 18 has the best

binding affinity followed by Peptide 8, 6, 15, 19, and 24. However, in MM/PBSA analysis, Peptide 18 (-5.4 ± 1.5 kcal/mol) was the third best in overall ranking following Peptide 15 (17.7 ± 1.8 kcal/mol) and 8 (8.7 ± 1.7 kcal/mol). The changes in ranking between docking and MM/PBSA analysis are due to the trajectory analysis and the behavior of the peptides inside the binding cavity throughout the simulation. Therefore, Peptide 18 showed a higher instability of binding to the active site compared to Peptides 15 and 8.

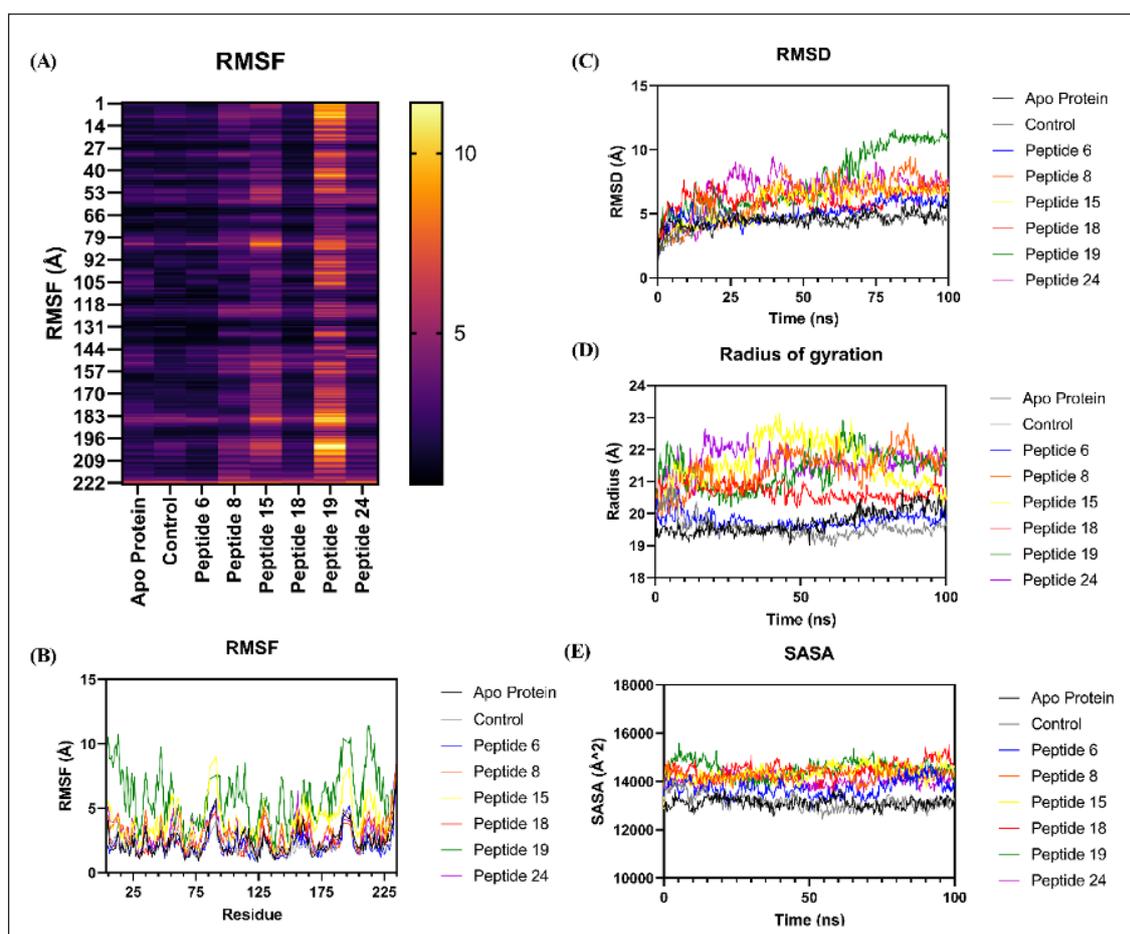
Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (Rg), and Solvent-Accessible Surface Area (SASA) analyses were performed over 100 ns MD simulations to obtain a quantitative assessment of the overall stability of Peptide 6, 8, 18, 19, 24 and C2E in complex with the MrkH protein. The root means square deviation (RMSD) for the six complexes studied was calculated to evaluate the conformational change and stability of the systems during the 100 ns MD simulation periods. Higher values of RMSD indicate relative instability, while lower RMSD values are more favorable. The RMSD plot for the six simulation systems can be seen in Figure 4.

As shown in Figure 4, Peptide 6 exhibited the highest stability with an RMSD value of 3.8 Å while Peptide 19 exhibited the lowest stability, i.e., a higher RMSD value (7.8 Å). The former peptide demonstrated better stability compared to the control ligand (4.4 Å), Peptide 15 (5.5 Å), 8 (5.8 Å), and 18 (5.8 Å). In terms of the binding affinity of the peptide to the active site, Peptide 6 performed better than the control ligand but not Peptide 15, 8, and 18 which have more positive values in MM/PBSA analysis. The lower RMSD values confer the stability of the complex throughout the simulation while the more positive value of MM/PBSA indicates a stable binding of the peptides to the MrkH.

Root-Mean-Square-Fluctuation (RMSF) was used to determine the flexibility and fluctuation of each MrkH residue and how much each residue moved throughout the simulation period. Each residue's flexibility was examined to understand better how peptide-binding affects protein flexibility during the MD simulation run. Lower values of RMSF mean better compactness, stiffness, and stability of the receptor. The RMSF values for the six complexes studied are shown in Figure 4a. Looking at the data in Figure 4b, it is noticeable that the RMSF results are compatible with the RMSD data. Peptides 19 and 15 showed relatively high RMSF values, indicating a high fluctuation of amino acid residues during complex formation with these inhibitors. It is worth noting that both peptide 8 and peptide 24 had closed RMSF values, i.e., a similar effect on amino acid residue fluctuation.

Table 6. Calculated the average MM/PBSA binding energies over 100 ns (1000 Snapshots) for the best peptides

Structure ID	MM/PBSA (kcal/mol)
Control	-46.1 ± 1.5
Peptide 6	-17.7 ± 1.8
Peptide 8	8.7 ± 1.7
Peptide 15	17.7 ± 1.8
Peptide 18	-5.4 ± 1.5
Peptide 19	-37.5 ± 2.1
Peptide 24	-16.0 ± 0.3

**Fig. 4.** Structural and conformational analysis for the best six peptides towards MrkH over 100 ns MD simulation. a) Root-mean-square deviation (RMSD). b) Radius of gyration (Rg). c) Solvent-accessible surface area (SASA). d) Root-mean-square fluctuation (RMSF).

The radius of Gyration (Rg) analysis was performed over 100 ns MD simulations to interpret the compactness of the protein structure in the system. Low Rg values indicate conformational stability and the degree to which the protein structure is densely packed. The Rg data are shown in Figure 4d. The behavior of peptide 6 concerning Rg was constant as in the data from SASA. From the beginning of 55 ns to the end of the MD simulation, peptide 18 showed very similar values to peptide 6. Moreover, apo-protein, C2E, and peptide 6 also showed remarkable overlap in Rg values from 20 ns to 100 ns.

Solvent-accessible surface area (SASA) is used to indicate the surface area of the receptor that is accessible to a solvent. Figure 4e shows the data from SASA for the six peptides, C2E, and the apoprotein. The estimated results show that peptide 6 has the lowest values of SASA compared to the others. On the other hand, peptide 19 showed a sharp increase in SASA values from 50 ns to the end of the MD simulation period. Both peptides 15 and 18 showed overlapping SASA values throughout the MD simulation period.

The majority of molecular interactions discovered during the MD simulations are consistent with the docking result. The plots in Figure 4 are very important because they show that both the peptides and the native ligand interact with the amino acids throughout the simulation and do not detach from their interaction site. However, the fluctuations in the RMSD and RMSF values of the peptides show that these peptides may be reorienting themselves during the simulation. Considering the MD simulation and the binding affinity of the peptides against the receptor, peptides 6 and 15 are suggested to be highly potential MrkH inhibitors compared to the rest of the peptides.

CONCLUSION

This study has shown that antimicrobial peptides derived from collagen were successfully predicted and analyzed using several ML platforms and *in silico* drug discovery tools. Further simulation of six potential antimicrobial cryptides using molecular dynamics approaches also demonstrated the cryptides-protein complex stability, validated by MM-PBSA analysis. All six cryptides are promising biological candidates to inhibit *K. pneumoniae* growth and biofilm synthesis. However, the cryptides have to be validated by experimental studies before they can be developed as potential therapeutic agents.

ACKNOWLEDGEMENTS

The study was funded by the Fundamental Research Grant Scheme (FRGS/1/2019/SKK11/UKM/03/1) from the Malaysian Higher Education Ministry while Siti Mariani Mhd Marzuki was sponsored by the Program Penyelidikan Prasiswazah Mutiara UKM (MUTIARA-A162762) from Universiti Kebangsaan Malaysia. We also acknowledge Associate Prof. Dr. Nazlina Ibrahim and Dr. Sylvia Chieng (Department of Biological Sciences and Biotechnology, Faculty of Sciences and Technology, Universiti Kebangsaan Malaysia) for their insight on the project and manuscript. In addition, we are thankful to Dr. Hafiz Muzazammel Rehman (School of Biochemistry and Biotechnology, University of Punjab) for providing an additional workstation for the molecular dynamic simulation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Abdillahi, H.S., Verschaeve, L., Finnie, J.F. & Van Staden, J. 2012. Mutagenicity, antimutagenicity and cytotoxicity evaluation of South African Podocarpus species. *Journal of Ethnopharmacology*, 139(3): 728-738. <https://doi.org/10.1016/j.jep.2011.11.044>
- Abdillahi, S. M., Bober, M., Nordin, S., Hallgren, O., Baumgarten, M., Erjefält, J., Westergren-Thorsson, G., Bjermer, L., Riesbeck, K., Egesten, A., & Mörgelin, M. 2015. Collagen VI is upregulated in COPD and serves both as an adhesive target and a bactericidal barrier for *Moraxella catarrhalis*. *Journal of Innate Immunity*, 7(5): 506-517. <https://doi.org/10.1159/000381213>
- Abdillahi, S. M., Maaß, T., Kasetty, G., Strömstedt, A. A., Baumgarten, M., Tati, R., Nordin, S. L., Walse, B., Wagener, R., Schmidtchen, A., & Mörgelin, M. 2018. Collagen VI contains multiple host defense peptides with potent in vivo activity. *The Journal of Immunology*, 201(3): 1007-1020. <https://doi.org/10.4049/jimmunol.1700602>
- Adekoya, O. A., & Sylte, I. 2009. The Thermolysin Family (M4) of Enzymes: Therapeutic and biotechnological potential. *Chemical Biology & Drug Design*, 73(1): 7-16. <https://doi.org/10.1111/j.1747-0285.2008.00757.x>

- Atef, M., Chait, Y.A., Ojagh, S.M., Latifi, A.M., Esmaili, M., Hammami, R. & Udenigwe, C.C. 2021. Anti-Salmonella activity and peptidomic profiling of peptide fractions produced from sturgeon fish skin collagen (*Huso huso*) using commercial enzymes. *Nutrients*, 13(8): 2657. <https://doi.org/10.3390/nu13082657>
- Baehaki, A., Suhartono, M. T., Sukarno, Syah, D., & Setyahadi, S. 2016. Collagen peptides from fish skin with Angiotensin I-Converting Enzyme (ACE) inhibitor and cancer antiproliferative activity. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 7(1): 1994-2000.
- Bahar, A.A. & Ren, D. 2013. Antimicrobial peptides. *Pharmaceuticals*, 6(12): 1543-1575. <https://doi.org/10.3390/ph6121543>
- Banerjee, P., & Shanthi, C. 2016. Cryptic Peptides from Collagen: A Critical Review. *Protein and Peptide Letters*, 23(7): 664-672. <https://doi.org/10.2174/0929866522666160512151313>
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. DiNola, A. & Haak, J.R. 1984. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8): 3684-3690. <https://doi.org/10.1063/1.448118>
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Research*, 28(1): 235-242. <https://doi.org/10.1093/nar/28.1.235>
- Blatt, J. M. & Weisskopf, V.F. 1979. Theoretical nuclear physics. In: *Theoretical Nuclear Physics* (1st Ed.). Springer New York. https://doi.org/10.1007/978-1-4612-9959-2_1
- Bradford, A., Raftery, M., Bowie, J., Tyler, M., Wallace, J., Adams, G. & Severini, C. 1996. Novel upeirin peptides from the dorsal glands of the Australian floodplain toadlet *Uperoleia inundata*. *Australian Journal of Chemistry*, 49(4): 475-484. <https://doi.org/10.1071/CH9960475>
- Budagavi, D.P. & Chugh, A. 2018. Antibacterial properties of Latarcin 1 derived cell-penetrating peptides. *European Journal of Pharmaceutical Sciences*, 115: 43-49. <https://doi.org/10.1016/j.ejps.2018.01.015>
- Chen, Y., Guarnieri, M.T., Vasil, A.I., Vasil, M.L., Mant, C.T. & Hodges, R.S. 2007. Role of peptide hydrophobicity in the mechanism of action of alpha-helical antimicrobial peptides. *Antimicrobial Agents and Chemotherapy*, 51(4): 1398-1406. <https://doi.org/10.1128/AAC.00925-06>
- Chung, C.-R., Jhong, J.-H., Wang, Z., Chen, S., Wan, Y., Horng, J.-T. & Lee, T.-Y. 2020. Characterization and identification of natural antimicrobial peptides on different organisms. *International Journal of Molecular Sciences*, 21(3): 986. <https://doi.org/10.3390/ijms21030986>
- Conlon, J.M., Sonnevend, A., Davidson, C., Smith, D.D. & Nielsen, P.F. 2004. The ascaphins: A family of antimicrobial peptides from the skin secretions of the most primitive extant frog, *Ascaphus truei*. *Biochemical and Biophysical Research Communications*, 320(1): 170-175. <https://doi.org/10.1016/j.bbrc.2004.05.141>
- Dobson, A., Cotter, P.D., Ross, R.P. & Hill, C. 2012. Bacteriocin production: A probiotic trait? *Applied and Environmental Microbiology*, 78(1): 1-6. <https://doi.org/10.1128/AEM.05576-11>
- Dressler, P., Gehring, D., Zdzieblik, D., Oesser, S., Gollhofer, A. & König, D. 2018. Improvement of functional ankle properties following supplementation with specific collagen peptides in athletes with chronic ankle instability. *Journal of Bodywork and Movement Therapies*, 22(4): 858. <https://doi.org/10.1016/j.jbmt.2018.09.037>
- Elber, R., Ruymgaart, A.P. & Hess, B. 2011. SHAKE parallelization. *The European Physical Journal*, 200(1): 211-223. <https://doi.org/10.1140/epjst/e2011-01525-9>
- Ennaas, N., Hammami, R., Gomaa, A., Bédard, F., Biron, É., Subirade, M., Beaulieu, L., & Fliss, I. 2016. Collagencin, an antibacterial peptide from fish collagen: Activity, structure and interaction dynamics with membrane. *Biochemical and Biophysical Research Communications*, 473(2): 642-647. <https://doi.org/10.1016/j.bbrc.2016.03.121>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. & Bairoch, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook*. J.M. Walker (Ed.). Humana Press. pp. 571-607. <https://doi.org/10.1385/1-59259-890-0:571>
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery Consortium & Raghava, G.P.S. 2013. In silico approach for predicting toxicity of peptides and proteins. *PLoS One*, 8(9): e73957. <https://doi.org/10.1371/journal.pone.0073957>
- He, J., Luo, X., Jin, D., Wang, Y., & Zhang, T. 2018. Identification, recombinant expression, and characterization of LHG2, a novel antimicrobial peptide of *Lactobacillus casei* HZ1. *Molecules*, 23(9): 2246. <https://doi.org/10.3390/molecules23092246>
- Holton, T.A., Pollastri, G., Shields, D.C. & Mooney, C. 2013. CPPpred: prediction of cell penetrating peptides. *Bioinformatics*, 29(23): 3094-3096. <https://doi.org/10.1093/bioinformatics/btt518>

- Honorato, R.V., Koukos, P.I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A. & Bonvin, A.M.J.J. 2021. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Frontiers in Molecular Biosciences*, 8: 729513. <https://doi.org/10.3389/fmolb.2021.729513>
- Houri, A.J. & Mechler, A. 2020. Mechanism of action of the antimicrobial peptide Caerin1.1. *ChemistrySelect*, 5(20): 5895-5902. <https://doi.org/10.1002/slct.202000851>
- Hubbard, R.E. & Kamran Haider, M. 2010. Hydrogen Bonds in Proteins: Role and Strength. In eLS. Wiley. <https://doi.org/10.1002/9780470015902.a0003011.pub2>
- Izaguirre, J.A., Catarello, D.P., Wozniak, J.M. & Skeel, R.D. 2001. Langevin stabilization of molecular dynamics. *The Journal of Chemical Physics*, 114(5): 2090-2098. <https://doi.org/10.1063/1.1332996>
- Jiang, Z., Vasil, A.I., Hale, J.D., Hancock, R.E.W., Vasil, M.L. & Hodges, R.S. 2008. Effects of net charge and the number of positively charged residues on the biological activity of amphipathic alpha-helical cationic antimicrobial peptides. *Biopolymers*, 90(3): 369-383. <https://doi.org/10.1002/bip.20911>
- Korkmaz, B., Horwitz, M.S., Jenne, D.E. & Gauthier, F. 2010. Neutrophil elastase, proteinase 3, and cathepsin G as therapeutic targets in human diseases. *Pharmacological Reviews*, 62(4): 726-759. <https://doi.org/10.1124/pr.110.002733>
- Krieger, E. & Vriend, G. 2015. New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry*, 36(13): 996-1007. <https://doi.org/10.1002/jcc.23899>
- Krieger, E., Koraimann, G. & Vriend, G. 2002. Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins*, 47(3): 393-402. <https://doi.org/10.1002/prot.10104>
- Krieger, E., Nielsen, J.E., Spronk, C.A.E.M. & Vriend, G. 2006. Fast empirical pKa prediction by Ewald summation. *Journal of Molecular Graphics & Modelling*, 25(4): 481-486. <https://doi.org/10.1016/j.jmgn.2006.02.009>
- Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P. & Tufféry, P. 2016. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*, 44(W1): W449-54. <https://doi.org/10.1093/nar/gkw329>
- Land, H. & Humble, M.S. 2018. YASARA: A tool to obtain structural guidance in biocatalytic investigations. *Methods in Molecular Biology*, 1685: 43-67. https://doi.org/10.1007/978-1-4939-7366-8_4
- Lata, S., Sharma, B. & Raghava, G. 2007. Analysis and prediction of antibacterial peptides. *BMC Bioinformatics*, 8(1): 263. <https://doi.org/10.1186/1471-2105-8-263>
- Mahlapuu, M., Håkansson, J., Ringstad, L. & Björn, C. 2016. Antimicrobial Peptides: An emerging category of therapeutic agents. *Frontiers in Cellular and Infection Microbiology*, 6: 194. <https://doi.org/10.3389/fcimb.2016.00194>
- Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. & Simmerling, C. 2015. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8): 3696-3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- Mao, Y., Niu, S., Xu, X., Wang, J., Su, Y., Wu, Y. & Zhong, S. 2013. The effect of an adding histidine on biological activity and stability of pc-pis from *Pseudosciaena crocea*. *PLoS ONE*, 8(12): e83268. <https://doi.org/10.1371/journal.pone.0083268>
- Martins, P.M., Santos, L.H., Mariano, D., Queiroz, F.C., Bastos, L.L., Gomes, I. de S., Fischer, P.H. C., Rocha, R.E.O., Silveira, S. A., de Lima, L.H.F., de Magalhães, M.T.Q., Oliveira, M.G.A. & de Melo-Minardi, R. C. 2021. Propedia: A database for protein-peptide identification based on a hybrid clustering algorithm. *BMC Bioinformatics*, 22(1): 1. <https://doi.org/10.1186/s12859-020-03881-z>
- Mobarki, N., Almerabi, B. & Hattan, A. 2019. Antibiotic resistance crisis. *International Journal of Medicine in Developing Countries*, 3(6): 561-564. <https://doi.org/10.24911/IJMDC.51-1549060699>
- Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B.H., Kumaran, E.A.P., McManigal, B., ... Naghavi, M. 2022. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*, 399(10325): 629-655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- O'Neill, J. 2016. Tackling drug-resistant infections globally: Final report and recommendations. London: HM Government and Wellcome Trust. Review on Antimicrobial Resistance, chaired by Jim O'Neill. https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf
- Osorio, D., Rondón-Villarreal, P. & Torres, R. 2015. Peptides: A package for data mining of antimicrobial peptides. *The R Journal*, 7(1): 4. <https://doi.org/10.32614/RJ-2015-001>

- Pan, Y.A., Xiao, X. & Wang, P. 2012. AMPpred: An on-line predictor design for automated AMP recognition. *Applied Mechanics and Materials*, 229-231: 2276-2279. <https://doi.org/10.4028/www.scientific.net/AMM.229-231.2276>
- Pirtskhalava, M., Amstrong, A.A., Grigolava, M., Chubinidze, M., Alimbarashvili, E., Vishnepolsky, B., Gabrielian, A., Rosenthal, A., Hurt, D.E. & Tartakovsky, M. 2021. DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Research*, 49(D1): D288-D297. <https://doi.org/10.1093/nar/gkaa991>
- Porto, W.F., Silva, O.N. & Franco, O.L. 2012. Prediction and rational design of antimicrobial peptides. In *Protein Structure*. E. Faraggi (Ed.). IntechOpen, London.
- Praet, S.F.E., Purdam, C.R., Welvaert, M., Vlahovich, N., Lovell, G., Burke, L.M., Gaida, J. E., Manzanero, S., Hughes, D. & Waddington, G. 2019. Oral supplementation of specific collagen peptides combined with calf-strengthening exercises enhances function and reduces pain in achilles tendinopathy patients. *Nutrients*, 11(1): 76. <https://doi.org/10.3390/nu11010076>
- Qutb, A. M., Wei, F. & Dong, W. 2020. Prediction and characterization of cationic arginine-rich plant antimicrobial peptide SM-985 from teosinte (*Zea mays* ssp. *mexicana*). *Frontiers in Microbiology*, 11: 1353. <https://doi.org/10.3389/fmicb.2020.01353>
- Roy, R., Tiwari, M., Donelli, G. & Tiwari, V. 2018. Strategies for combating bacterial biofilms: A focus on anti-biofilm agents and their mechanisms of action. *Virulence*, 9(1): 522-554. <https://doi.org/10.1080/21505594.2017.1313372>
- Schumacher, M. A., & Zeng, W. 2016. Structures of the activator of *K. pneumoniae* biofilm formation, MrkH, indicates PilZ domains involved in c-di-GMP and DNA binding. *Proceedings of the National Academy of Sciences of the United States of America*, 113(36): 10067-10072. <https://doi.org/10.1073/pnas.1607503113>
- Sharma, A., Gupta, P., Kumar, R. & Bhardwaj, A. 2016. dPABBs: A Novel in silico approach for predicting and designing anti-biofilm peptides. *Scientific Reports*, 6: 21839. <https://doi.org/10.1038/srep21839>
- Shi, G., Kang, X., Dong, F., Liu, Y., Zhu, N., Hu, Y., Xu, H., Lao, X. & Zheng, H. 2022. DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. *Nucleic Acids Research*, 50(D1): D488-D496. <https://doi.org/10.1093/nar/gkab651>
- Struve, C., Bojer, M., & Krogfelt, K.A. 2009. Identification of a conserved chromosomal region encoding *Klebsiella pneumoniae* type 1 and type 3 fimbriae and assessment of the role of fimbriae in pathogenicity. *Infection and Immunity*, 77(11): 5016-5024. <https://doi.org/10.1128/IAI.00585-09>
- Swanson, J.M.J., Henchman, R.H. & McCammon, J.A. 2004. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophysical Journal*, 86(1): 67-74. [https://doi.org/10.1016/S0006-3495\(04\)74084-9](https://doi.org/10.1016/S0006-3495(04)74084-9)
- Torrent, M., Andreu, D., Nogués, V.M. & Boix, E. 2011. Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PLoS One*, 6(2): e16968. <https://doi.org/10.1371/journal.pone.0016968>
- van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J. & Bonvin, A.M.J.J. 2016. The HADDOCK2.2 Web server: user-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, 428(4): 720-725. <https://doi.org/10.1016/j.jmb.2015.09.014>
- Vangone, A., Schaarschmidt, J., Koukos, P., Geng, C., Citro, N., Trellet, M. E., Xue, L.C. & Bonvin, A.M.J.J. 2019. Large-scale prediction of binding affinity in protein-small ligand complexes: the PRODIGY-LIG web server. *Bioinformatics*, 35(9): 1585-1587. <https://doi.org/10.1093/bioinformatics/bty816>
- Waghu, F.H., Barai, R.S., Gurung, P. & Idicula-Thomas, S. 2016. CAMPR3: A database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Research*, 44(D1): D1094-D1097. <https://doi.org/10.1093/nar/gkv1051>
- Wilksch, J.J., Yang, J., Clements, A., Gabbe, J.L., Short, K.R., Cao, H., Cavaliere, R., James, C.E., Whitchurch, C.B., Schembri, M.A., Chuah, M.L.C., Liang, Z.-X., Wijburg, O.L., Jenney, A.W., Lithgow, T. & Strugnell, R.A. 2011. MrkH, a novel c-di-GMP-dependent transcriptional activator, controls *Klebsiella pneumoniae* biofilm formation by regulating type 3 fimbriae expression. *PLoS Pathogens*, 7(8): e1002204. <https://doi.org/10.1371/journal.ppat.1002204>
- Win, T.S., Malik, A.A., Prachayasittikul, V., Wikberg, J.E.S., Nantasenamat, C. & Shoombuatong, W. 2017. HemoPred: a web server for predicting the hemolytic activity of peptides. *Future Medicinal Chemistry*, 9(3): 275-291. <https://doi.org/10.4155/fmc-2016-0188>

- Yeaman, M.R. & Yount, N.Y. 2003. Mechanisms of antimicrobial peptide action and resistance. *Pharmacological Reviews*, 55(1): 27-55. <https://doi.org/10.1124/pr.55.1.2>
- Zdzieblik, D., Brame, J., Oesser, S., Gollhofer, A. & König, D. 2021. The influence of specific bioactive collagen peptides on knee joint discomfort in young physically active adults: A randomized controlled trial. *Nutrients*, 13(2): 523. <https://doi.org/10.3390/nu13020523>
- Zheng, J., Lin, Z., Chen, C., Chen, Z., Lin, F., Wu, Y., Yang, S., Sun, X., Yao, W., Li, D., Yu, Z., Jin, J., Qu, D. & Deng, Q. 2018. Biofilm formation in *Klebsiella pneumoniae* bacteremia strains was found to be associated with CC23 and the presence of wcaG. *Frontiers in Cellular and Infection Microbiology*, 8: 21. <https://doi.org/10.3389/fcimb.2018.00021>

