



Insilico Design, Evaluation Of Novel Antimicrobial Peptide SS-BF-36 Targeting Biofilms And Multidrug-Resistant Organisms

Sivaa Arumugam Ramakrishnan¹, Sindhu Kalajirao², Alin Bose J³, Vasanth Raj Palanimuthu⁴, Raman Rajeshkumar^{5*}

Abstract

In the global pandemic, several micro-organisms began to become resistant to the drugs due to abuse or prolonged use. They also get multidrug resistance through quorum sensing. The modulation of gene expression in response to oscillations in cell-population density is known as quorum sensing. With quorum sensing, the benefit increases as the cell population increases. Biofilms consist of a large number of extracellular polymer compounds that are released by the bacteria living in them. By ingesting this extracellular polymeric substance, they become cross-resistant to the given drugs and act as a protective layer for the pathogenic microbes. To overcome these biofilms, antimicrobial peptides (AMP) can be characterized, insilico rather using classical techniques. Several insilico techniques and machine learning techniques were employed to identify an unidentified effective AMP against the biofilms produced by *P. aeruginosa* mixed culture. In machine learning, Random Forest algorithm with decision tree was used to predict the AMP sequences. The predicted AMP sequence was combined with breakable linkers and cytotoxicity against a human was studied. It is a preliminary study that can be further investigated in the future to target biofilms against resistant micro-organisms.

1655

Keywords: AMP, Biofilms, Quorum Sensing, insilico, Machine Learning.

DOI Number: 10.48047/NQ.2022.20.17.NQ880203

Neuroquantology 2022; 20(17):1655-1660

1. INTRODUCTION

The threat of the upcoming future is microbial resistance rather than newly evolving pathogenic microorganism. Antimicrobial resistance increases as the time of drug usage or misuse of drugs increases. So, if all drugs for a particular pathogenic microorganism get resisted, it becomes harder than the current situation which is happening now, for example, Multi-Drug Resisted (MDR) tuberculosis. These resistances occur because of quorum sensing, and biofilms are formed to protect these microorganisms further from antibiotics. These biofilms are mostly harmful and form in many medical devices such as mechanical heart valves, breast implants, urine catheters (Preda VG, Săndulescu O, 2019). It doesn't form biofilm

good environment for pathogenic micro-organisms to get resisted. It can be prevented by discovering novel naturally occurring antimicrobial peptides. To overcome the traditional peptides, AMPs are designed as it has lower toxicity and production cost is less comparatively.

2. Methodology

2.1. Analysis of the Amino acid sequence of AMP

Antimicrobial peptides should have a sequence length of a minimum of 10 and a maximum of 29 and they have several types in them. In this study, an anti-biofilm peptide was designed *insilico* and evaluated for its properties. Previously identified a novel AMP Chrysopsin-

***Corresponding Address:** - Raman Rajeshkumar

Address: ^{1,2,3,4}Department Of Pharmaceutical Biotechnology, JSS College Of Pharmacy, JSS Academy Of Higher Education & Research, Ooty, Nilgiris, Tamil Nadu, India.

⁵JSS College Of Pharmacy Rocklands, Post Box No.20, Udhagamandalam - 643 001 Tamil Nadu, India, Email: bathmic@jssuni.edu.in

Relevant Conflicts Of Interest/Financial Disclosures: The Authors Declare That The Research Was Conducted In The Absence Of Any Commercial Or Financial Relationships That Could Be Construed As A Potential Conflict Of.



1 (Wei Wang et al, 2012) was selected and the amino acid sequence of the peptide was taken. Chrysophsin-1 was identified from the fish, and it has a sequence length of 25 amino acids with molecular weight - 2890.662, and isoelectric point - 12.813. Then this sequence is retrieved from BaAMPs (<http://www.baamps.it/>) (M. Di Luca, 2015) and further studied.

2.2. BLASTp of sequence.

The sequence of the amino acid (AMP-Chrysophsin-1) was procured and the sequence was blasted in BLASTp (RRID:SCR_001010) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) and from the results, only the microorganism was selected for easy production and reproducibility. The organism chosen was *Pseudomonas stutzeri* and type VI secretion system tip protein VgrG, partial (NCBI Reference Sequence: WP_044315960.1) was taken and the protein sequence was elucidated for the antimicrobial properties.

2.3. Data curation for machine learning.

The antimicrobial peptide sequence database was curated from the BaAMPs (<http://www.baamps.it/>), and positive and negative datasets were meticulously curated to form the dataset. All the sequence of the dataset was converted to the AAC by the self-written software. Then the machine learning software was programmed with the RandomForest algorithm, additional useful modules, and packages to build an ML. The ML was trained with both positive and negative datasets. Then the sequence to be tested was given as the input. The software itself converts the sequence, into AAC and compares it with the trained dataset (<https://doi.org/10.5281/zenodo.3509134>) to get the predictive results. The predictive result of the sequence will be shown as a graph in the software.

2.4. Prediction of AMP activity.

The complete protein sequence was predicted for the antimicrobial peptides in it. It was predicted with machine learning (ML). It also includes three types of machine learning i.e., Support Vector Machine algorithm (SVM) and WEKA algorithm in dPABBs (<https://ab-openlab.csir.res.in/abp/antibiofilm/protein.php>) (Gupta, P. et al, 2016) and Random Forest (Own program). Comparative to the SVM, WEKA, Random Forest has well defined visual and predictive model. So, WEKA is used to determine the AMP in the protein sequence in it. From the predictive peptide sequences, it is

further performed 3-fold with SVM, WEKA, and Random Forest algorithms. Common predicted peptide sequences were procured from predicted peptide sequences in both SVM, WEKA, and Random Forest algorithms were only taken, and the rest is omitted for the most accurate results.

2.5. Toxicity study of AMP sequences

The predicted peptide sequences were subjected to Toxinpred (Gupta, P. et al, 2013) (<https://webs.iitd.edu.in/raghava/toxinpred/design.php>) to determine toxicity for humans. The reason to test the cytotoxicity for humans is to protect the human from any accidental intake of the AMP but should be toxic to the microorganisms. So, the non-toxic peptide sequences with AMP activity were scrutinized and proceeded to build one modified sequence.

2.6. Building AMP sequence.

Predicted peptide sequences in SVM, WEKA, and Random Forest were taken and these peptide sequences were joined together with breakable linkers (Chen X, Zaro JL & Shen WC, 2013). An extreme short peptide WWW was also joined with the modified AMP sequence to increase its stability. These joined peptide sequences were further predicted with the Random Forest algorithm. Thus, the sequence will undergo a 4-fold evaluation. These predicted peptide sequences were linked together and were further elucidated for the structural analysis.

2.7. 3D Structural prediction of AMP sequence and elucidation

The 3D structure of the modified antibiofilm peptide sequence was predicted with the RaptorX (Morten Källberg et al, 2012) (<http://raptorx.uchicago.edu/ContactMap/>) and the structure was collected in the PDB format and RCSB PDB viewer (D. Sehnal et al, 2021) (<https://www.rcsb.org/3d-view>) was used to see the PDB file. Protparam web server (Gasteiger E et al, 2005) (<https://web.expasy.org/protparam/>) was used to elucidate the number of amino acids present in the sequence, the molecular weight of the sequence containing amino acids, theoretical PI of the sequence, the total number of negatively charged residues present in the modified sequence, the total number of positively charged residues present in the modified sequence, atomic composition of the modified sequence, and sequence chemical formula, the total number of atoms present in the modified sequence, aliphatic index of the modified



sequence and index of instability with hydrophobicity grand average. Predicted AMPs were subjected to the pre-trained dataset to further scrutinize the activity which increases the evaluation into 5 folds. These were predicted with the webserver iAMPpred (Meher P.K et al, 2017) (<http://cabgrid.res.in:8080/amppred/server.php>) to determine the AMPs antibacterial, antiviral and antifungal activities. The predictions were calculated as individual AMPs, as they will cleave into individuals at the site of action owing to the cleavable rigid linker.

3. Results

3.1. Analysis of chrysophin-1

Chrysophin-1 was the antimicrobial peptide having a peptide length of 25 was selected for this study. Though it has AMP activity, it can't be considered as the novel AMP as the chrysophin-1 is an unstable peptide with a higher hydrophobic moiety. This became a cause to identify new novel stable AMP. The sequence of the chrysophin-1 was taken for the next stage in identifying new AMP. The properties of chrysophin-1 are mentioned in Table 1.

Table 1: Shows the properties of the chrysophin-1.

| Name | Sequence | Size | AMP Source |
|------------------------|---------------------------|------|------------|
| Chrysophin-1 | FFGWLIKGAIHAGKAIHGLIHRRRH | 25 | Fish |
| Name | Value | | |
| NetCharge@5 | 8.915 | | |
| NetCharge@7 | 5.937 | | |
| Isoelectric Point | 12.813 | | |
| Molecular Weight | 2890.662 | | |
| Extinction Coefficient | 5500 | | |
| Hydrophobicity (CCS) | -0.152 | | |
| Hydrophobic Mom (CCS) | 2.345 | | |

3.2. Discovery of a similar sequence using BLASTp.

The chrysophin-1 sequence was taken and BLASTp (protein blasted) in the NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) to get similar types of peptide sequences. Since many of the similar peptides were from the fish, some of the peptides were from different organisms. A protein sequence of *Pseudomonas stutzeri* (Type VI secretion system tip protein VgrG, partial, NCBI Reference Sequence: WP_044315960.1) with an average above 60 percent similarity was selected and the sequence was used to predict the peptide sequences with AMP activity in the upcoming process.

3.3. Preparation of dataset for ML

On another side, the AMP sequence dataset was obtained from the BaAMPs

(<http://www.baamps.it/>). It was classified into a positive dataset and negative dataset from a minimum peptide length of 10. A total of 187 peptide sequences were used and the AAC was calculated for each sequence. Then these features were converted into arrays to create a database for machine learning.

3.4. Algorithms predicted the AMP activity of the sequence.

From the protein sequence, the AMPs were identified with the assistance of the dPABBs server with the whole amino acid composition in SVM based model, and the active AMP's in the SVM model was collected and further evaluated in the WEKA based model. Most of the peptide sequences are inactive against biofilms and only the peptide, active against the biofilm proceed from SVM to the WEKA algorithm. The AMP's which are active in both models have proceeded to the Random Forest algorithm. The above-scrutinized peptides which have activity in Random Forest are been marked for the next stage of the process (Table 2 & Figure 1)

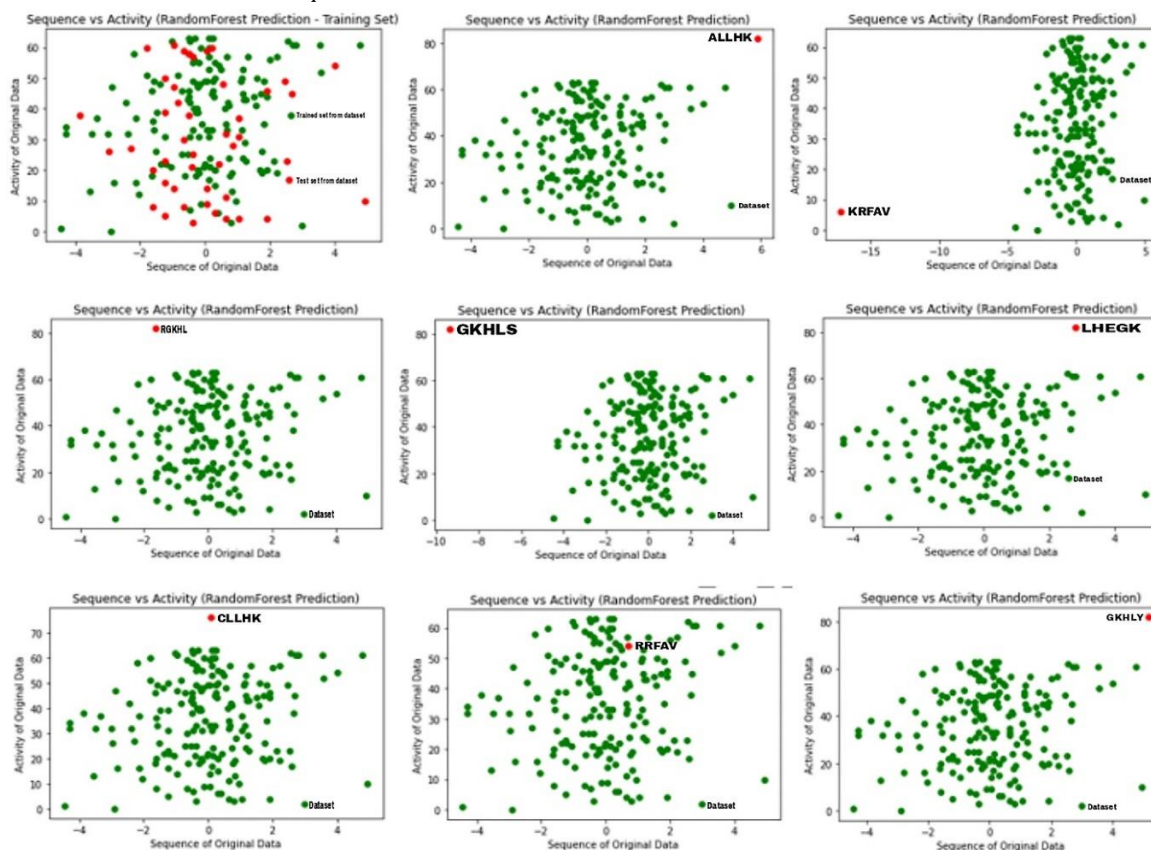
Table 2: Prediction of the peptide activity by SVM, WEKA, and Random

| Sequence | SVM | WEKA | Random forest | Selected / Rejected |
|----------|-----|------|---------------|---------------------|
| ALLHK | Yes | No | Yes | Rejected |
| KRFAV | Yes | No | No | Rejected |
| RGKHL | Yes | Yes | Yes | Selected |
| GKHLs | Yes | No | No | Rejected |
| LHEGK | Yes | No | Yes | Rejected |
| CLLHK | Yes | Yes | Yes | Selected |
| RRFAV | Yes | Yes | Yes | Selected |
| GKHLY | Yes | Yes | Yes | Selected |

In the first image, the mixed dataset was separated into two groups as a training set and the next one as a sample set. In the training set, the number of positive sequences was much greater (ratio = 20:1) to the negative sequences and the sample set contains a random number of the positive set & sample set. In the Random Forest graph, the closest distance of the sample to the trained dataset shows the prediction rate, activity is shown by the y-axis, and AAC is shown by the x-axis. In this ALLHK, RGKHL, LHEGK, CLLHK, RRFAV, and GKHLy have the nearest distance to the training set, and GKHLs, KRFAV has an extreme distance to the training set. Thus, these two are predicted as inactive peptides by the Random Forest algorithm.



Figure 1: Shows the activity prediction of the peptides with the Random Forest algorithm. The red dot shows the sample, and the green dot shows the trained sequence.



3.5. Conjecturing toxicity of AMP sequences.

The algorithms predicted individual AMP sequences of the peptide length 5 was evaluated for its toxicity with the webserver toxinpred (<https://webs.iitd.edu.in/raghava/toxinpred/design.php>). It uses SVM-based prediction to predict the toxicity of the AMP-predicted sequences. The prediction of the selected AMP's was given in Table 3.

Table 3: Shows the toxicity of the selected AMP's and only nontoxic AMPs were selected to intercept any toxic effects to humans.

| Sequence | Toxicity |
|----------|-----------|
| RGKHL | Non-toxic |
| CLLHK | Non-toxic |
| RRFAV | Non-toxic |
| GKHLV | Non-toxic |

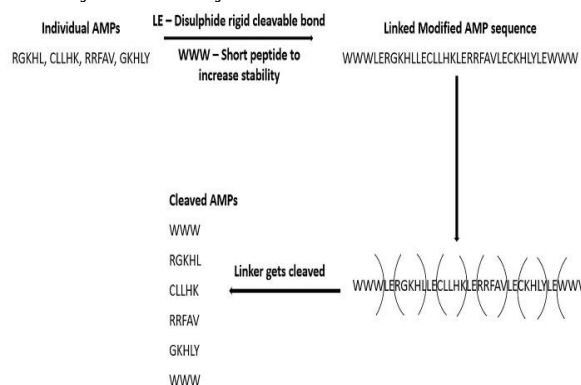
3.6. Modified AMP sequence.

The nontoxic AMPs were built into one sequence with a rigid disulfide cleavable bond. The rigid cleavable bond is used as the rigidity doesn't allow to express the side ends to react

and the linker cleaves to release the whole modified AMP sequence into small AMPs of

peptide length 5. A small peptide was attached front and end of the sequence to increase the stability. It is clearly described in Figure 2 and the prediction of AMP activity of the modified sequence is shown in Figure 3.

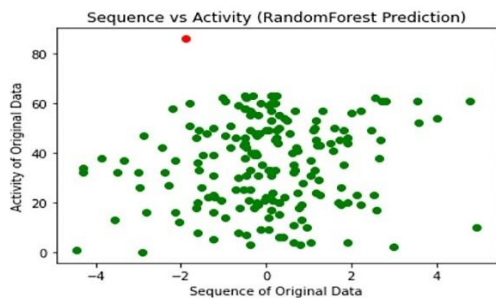
Figure 2: Explains the cleavage of modified AMP sequence to release individual AMP's which react on the resistant micro-organism and biofilms, but WWW peptide doesn't have any AMP activity.



Since the activity of the individual sequences was calculated and their mean value was fitted with the trained dataset.



Figure 3: Shows the AMP activity of the modified sequence by the Random Forest algorithm. The red dot shows the modified sequence, and the green dot shows the trained set of active AMPs.



3.7. 3D Structural prediction of AMP sequence and elucidation

RaptorX

(<http://raptorx.uchicago.edu/ContactMap/>)

was used to identify the 3D structure of the modified antibiofilm peptide sequence and the 3D structure was collected in the PDB format and viewed with RCSB PDB viewer

(<https://www.rcsb.org/3d-view>) which is shown in Figure 4. 2.8% of A, 8.3% of R, 2.8% of C, 13.9% of E, 5.6% of G, 8.3% of H, 25% of L, 8.3% of K, 2.8% of F, 16.7% of W, 2.8% of Y, 2.8% of V - AAC present in the sequence, 4761.62 - molecular weight of the sequence containing amino acids, 8.18 - theoretical PI of the sequence, 5 - total number of negatively charged residues present in the modified sequence, 6 - total number of positively charged residues present in the modified sequence, C - 232, H - 332, N - 60, O - 48, S - 1 - atomic composition of the modified sequence, and $C_{232}H_{332}N_{60}O_{48}S_1$ - sequence chemical formula, 673 - total number of atoms present in the modified sequence, 108.33 - aliphatic index of the modified sequence and 39.59% - index of instability with -0.397 - hydropathicity grand average were elucidated with Protparam webserver

(<https://web.expasy.org/protparam/>). The stability comparison between the chrysophin-1 and newly discovered is given in Table 4 Predicted AMPs were evaluated with a pre-trained dataset in iAMPpred to determine the activity of the individual peptide sequence since the linkers will get cleaved and released as the separate AMPs. The results were mentioned in Table 5.

Table 4: Shows the stability comparison between chrysophin-1 and newly discovered modified sequence (SS-BF-36).

| Chrysophin-1 | SS-BF-36 |
|--------------|----------|
| Unstable | Stable |

Figure 4: The modified sequence was converted into a structure with RaptorX and visualized with RCSB PDB Viewer.

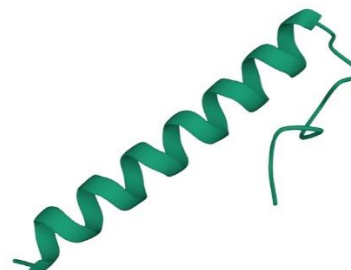


Table 5: The AMPs were predicted with the iAMPpred to evaluate its antibacterial, antiviral, and antifungal activities.

| Sequence | Antibacterial | Antiviral | Antifungal |
|----------|---------------|-----------|------------|
| RGKHL | 0.83 | 0.64 | 0.83 |
| CLLHK | 0.75 | 0.67 | 0.52 |
| RRFAV | 0.52 | 0.27 | 0.24 |
| GKHL | 0.79 | 0.67 | 0.80 |

Conclusion

Several drugs are present in the science community. They may be found earlier or recently but most of the less toxic drugs are being omitted to treat the organisms. This is not due to the decrease in the efficiency of the drug but an increase in the resistance of the drug by the microorganisms. One of the main causes for the increasing resistance in the microorganisms in the production of biofilms, a protective layer, and quorum sensing. Due to the toxic chemicals, it is difficult to wash the catheters, pacemakers, and other internal instruments before or after placing the surgery. So, it is necessary to develop nontoxic AMPs to treat the biofilm even while the medical instruments are inside the body. It is necessary to develop an AMP which is more stable than chrysophin-1. By the above evaluation of the SS-BF-36, it is more stable than chrysophin-1 and has an efficient AMP activity. Since it is a stable sequence, the formulations can be prepared with ease compared to the chrysophin-1. SS-BF-36 formulation can be used even on open wounds to prevent the infection caused by the *P. aeruginosa* with the above-predicted results in *insilico*. So, we conclude that SS-BF-36 is a novel AMP as it is stable with good



AMP activity, and it can be performed in *in vivo* and *in vitro* in the future.

Author Contribution

Sivaa Arumugam R*: Formal analysis, Methodology, Investigation, Data Curation, Writing - Original Draft, Writing -Review & Editing, ML Programming, and Visualization. Sindhu K*: Formal analysis, Methodology, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Programming and Visualization (*Equal contribution as the first author), Alin Bose J & Vasanth Raj Palanimuthu*: Data Curation & Review & Editing Raman Rajeshkumar*: Conceptualization, Methodology, Validation, Writing -Review & Editing, Supervision

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

Reference

- Preda VG, Săndulescu O. Communication is the key: biofilms, quorum sensing, formation and prevention. *Discoveries (Craiova)*. 2019 Sep 30;7(3):e100. doi: 10.15190/d.2019.13. PMID: 32309618; PMCID: PMC7086079.
- Wang W, Tao R, Tong Z, Ding Y, Kuang R, Zhai S, Liu J, Ni L. Effect of a novel antimicrobial peptide chrysopsin-1 on oral pathogens and *Streptococcus mutans* biofilms. *Peptides*. 2012 Feb;33(2):212-9. doi: 10.1016/j.peptides.2012.01.006. Epub 2012 Jan 20. PMID: 22281025.
- M. Di Luca, G. Maccari, G. Maisetta, G. Batoni BaAMPs: the database of biofilm-active antimicrobial peptides *Biofouling* 2015; 31(2):193-9.
BLASTP (RRID:SCR_001010)
<https://doi.org/10.5281/zenodo.3509134>
- Gupta, P. et al. dPABBs: A Novel in silico Approach for Predicting and Designing Anti-biofilm Peptides. *Sci. Rep.* 6, 21839; doi: 10.1038/srep21839 (2016).
- Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP, Open Source Drug Discovery Consortium (2013) In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE* 8(9):e73957
- Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design, and functionality. *Advanced drug delivery reviews*. 2013 Oct 15;65(10):1357-69.
- Morten Källberg, Haipeng Wang, Sheng Wang, Jian Peng, Zhiyong Wang, Hui Lu & Jinbo Xu. Template-based protein structure modeling using the RaptorX web server. *Nature Protocols* 7, 1511–1522, 2012.
- D. Sehnal, S. Bittrich, M. Deshpande, R. Svobodová, K. Berka, V. Bazgier, S. Velankar, S.K. Burley, J. Koča, A.S. Rose (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular

structures. *Nucleic Acids Research*. doi: 10.1093/nar/gkab314

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press (2005).pp. 571-607

Meher, P. K., Sahu, T. K., Saini, V. and Rao, A. R. (2017). Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical, and structural features into Chou's general PseAAC. *Scientific Reports*, 7(1), 1-12.

