1 DATA AVAILABILITY
The datasets generated and analyzed in this study, along with codes, can be found in the Antibiofilm repository at github.com/davidanastasiu/antibiofilm.

2 DATASET STATISTICS
Table S1 presents the number of peptides for training, validation and out-of-sample test sets for both the positive and negative datasets. The table also contains details of the dataset used for training and evaluating the regression models.

Table S1. Dataset Distribution of Our Machine Learning Models

| Dataset          | Type     | Training | Validation | Out-of-Sample Test |
|------------------|----------|----------|------------|--------------------|
| Classification   | Positive | 175      | 19         | 48                 |
|                  | Negative | 1741     | 194        | 485                |
| MBIC Classification | ≤ 64µM  | 128      | 32         | N/A                |
|                  | > 64µM   | 14       | 4          | N/A                |
| MBIC Regression  | ≤ 64µM   | 128      | 32         | N/A                |
|                  | ≤ 64µM   | 33       | 9          | N/A                |
| Candidate        |          | 135015   |            |                    |

3 CHARACTERIZATION OF PEPTIDES
Figure S1 presents the ten dipeptides with the highest composition percentage from the negative dataset. Interestingly, most of the dipeptides in the top ten set contain leucine, a non-polar amino acid.
Figure S1. Dipeptide composition of the negative dataset: all the dipeptides contain the non-polar amino acid leucine;

4 PERFORMANCE OF MACHINE LEARNING MODELS

Table S2 presents results from our evaluation of different machine learning models based on individual features while Table S3 displays the performance of different models when we combine two features together. Finally, Table S4 showcases the performance of our models when we combine more than two features. Our best performing model combines the AAC, DPC, CTD and Motif features.

Table S2. Performance Evaluation of Different Machine Learning Techniques with Individual Features

| Feature | Model  | Sensitivity | Specificity | Accuracy | F1 Score | MCC   |
|---------|--------|-------------|-------------|----------|----------|-------|
| AAC     | SVM    | 72.91       | 99.79       | 97.37    | 83.33    | 82.93 |
|         | RF     | 68.75       | 100         | 97.18    | 81.48    | 81.66 |
|         | XGBoost| 75.11       | 99.38       | 97.18    | 82.75    | 81.76 |
| DPC     | SVM    | 85.41       | 98.35       | 97.17    | 84.35    | 82.99 |
|         | RF     | 72.91       | 99.17       | 96.81    | 80.45    | 79.24 |
|         | XGBoost| 79.16       | 98.76       | 96.99    | 82.60    | 81.06 |
| CTD     | SVM    | 83.33       | 99.38       | 97.93    | 86.94    | 87.91 |
|         | RF     | 70.83       | 99.79       | 97.18    | 81.92    | 81.62 |
|         | XGBoost| 85.41       | 98.96       | 97.74    | 87.23    | 86.02 |
Table S3. Performance Evaluation of Different Machine Learning Techniques with a Combination of Two Features

| Features     | Model   | Sensitivity | Specificity | Accuracy | F1 Score | MCC    |
|--------------|---------|-------------|-------------|----------|----------|--------|
| AAC & DPC    | SVM     | 81.25       | 98.96       | 97.33    | 84.78    | 83.33  |
|              | RF      | 77.08       | 99.79       | 97.74    | 86.08    | 85.52  |
|              | XGBoost  | 81.25       | 99.38       | 97.74    | 86.66    | 85.76  |
| DPC & CTD    | SVM     | 77.08       | 99.79       | 97.74    | 86.04    | 85.52  |
|              | RF      | 70.83       | 100         | 97.37    | 82.92    | 82.97  |
|              | XGBoost  | 79.16       | 99.17       | 97.37    | 84.44    | 83.23  |
| CTD & AAC    | SVM     | 85.41       | 99.17       | 97.93    | 88.17    | 87.09  |
|              | RF      | 72.91       | 100         | 97.56    | 84.33    | 84.26  |
|              | XGBoost  | 79.16       | 99.58       | 97.74    | 86.36    | 85.56  |

Table S4. Performance Evaluation of Different Machine Learning Techniques with a Combination of Three or More Features

| Features                              | Model                             | Sensitivity | Specificity | Accuracy | F1 Score | MCC    |
|---------------------------------------|-----------------------------------|-------------|-------------|----------|----------|--------|
| AAC & DPC & CTD                       | SVM (c=100, gamma=0.01)          | 85.41       | 98.96       | 97.94    | 88.42    | 87.29  |
|                                       | RF (n-estimator=100)              | 70.83       | 100         | 97.37    | 82.92    | 82.97  |
|                                       | XGBoost (n-estimator=100, gamma=0.5) | 81.25       | 99.58       | 97.93    | 87.64    | 86.84  |
| AAC & DPC & CTD & motif               | SVM (c=150, gamma=0.05, Motif=ALL) | **85.48**   | **99.79**   | **98.49** | **91.11** | **90.53** |
|                                       | RF (with Motif=BETTS-RUSSELL)     | 72.91       | 100         | 97.56    | 84.33    | 84.26  |
|                                       | XGBoost (with Motif=BETTS-RUSSELL) | 85.41       | 99.38       | 98.12    | 89.13    | 88.20  |

Table S5. Performance Comparison of Our Method with the Dataset from Gupta et al. (2016)

| Validation dataset performance       | Specificity | Sensitivity | Accuracy | F1 Score | MCC    |
|--------------------------------------|-------------|-------------|----------|----------|--------|
| Reported in Gupta et al. (2016)      | 97.75       | 91.67       | 97.19    | N/A      | 0.84   |
| Achieved with our model              | 99.71       | 86.11       | **98.46** | **91.17** | **0.90** |
5 CHARACTERIZATION OF PEPTIDES FROM THE MBEC DATASET

We present the characteristics of the 57 peptides that were selected for training the regression model responsible for predicting the MBEC value of a candidate antibiofilm peptide.

![Figure S2](image_url)

Figure S2. Performance and characteristics of peptides with MBIC/MBEC values; (A) Number of peptides with MBEC values in different sequence ranges; (B) Percentage of helices, strands and coils in secondary structure for peptides with MBEC values;

6 NEWLY FOUND ANTIBIOFILM PEPTIDES

6.1 Visualization

We have evaluated the 2D structures of the peptides using PEP2D server Singh et al. (2019).

We further evaluated the structure of the peptides with probable antibiofilm activity. We evaluated helical wheel structure (Figure S5) for the peptides which showed higher percentage of helices in secondary structure evaluation.
Figure S3. Predicted 2D structures of previously characterized peptides with potential antibiofilm activity.

The 2D structures were evaluated using the PEP2D server. The pink cylinders represent helix, yellow arrows represent sheet, and the black line is coil.
Supplementary Material

Figure S4. Predicted 2D structures of previously characterized peptides with potential antibiofilm activity.

The 2D structures were evaluated using the PEP2D server. The pink cylinders represent helix, yellow arrows represent sheet, and the black line is coil.
Figure S5. The helical wheel structures of a few newly found antibiofilm peptides; (A) P17238, Mastoparan; (B) C0HK43, Lasioglossin; Rest of the peptide marked as ‘uncharacterized protein’ (C) A0A2P2Q2Y8, (D) A0A0A9U210, (E) E9JAR4, (F) A0A3Q7GQZ6. Here, hydrophilic amino acids are shown in circles, hydrophobic as diamonds. Negatively charged amino acids are triangles, and positively charged are pentagons. The hydrophobic amino acids are green, and the green shade decreases to yellow as per decreasing hydrophobicity. Hydrophilic amino acids are in red and the amount of red decreases as per decreasing hydrophilicity. The highly charged amino acids are in light blue and non-polar amino acids are in dark red. The numbers indicate the hydrophobic moment and the direction of the moment. The wheel structures were obtained using the software created by Don Armstrong and Raphael Zidovetzki, version 1.4, 2009-10-20 Schiffer and Edmundson (1967); Armstrong and Zidovetzki (2009).
6.2 Alignment

We also analyzed a few newly found antibiofilm peptides against some well known antibiofilm peptides which already have an eradication effect on preformed biofilm. For example, we aligned human cathelicidin, LL-37, against the set of Mastoparan-like peptides from our list. The alignment was done using the Clustal default webservice Madeira et al. (2019). The alignment is displayed in Figure S6 using Jalview V2 Waterhouse et al. (2009).

Figure S6. Pairwise sequence alignment of peptide hits with their closely matching known antibiofilm peptide. The default colour scheme used as per ClustalX

Colorcode – blue: residue A, I, L, M, F, W, V; red: residue K, R; green: residue N, Q, S, T; magenta: E, D; yellow: residue P.
6.3 Peptide List

The list of probable antibiofilm peptides from our pipeline are listed in Tables S6–S13. The tables contain peptide sequences and predicted MBEC values. We grouped the peptides in several MBIC value ranges.

Table S6. Newly Predicted Antibiofilm Peptides with MBIC Range 1–8 (µM) from the DRAMP database

| Name             | Seq                                     | Source     | Predicted MBEC (µM) |
|------------------|-----------------------------------------|------------|---------------------|
| DRAMP04642       | GIGKFLHSAGKFGKAFGEIMKS                  | Synthetic  | 1.510               |
| DRAMP02364       | GFWGKFLKLGIIHIGILHLLHL                 | Mammals    | 1.617               |
| DRAMP04663       | ARRLKLKAKKILKWL                        | Sheep      | 1.617               |
| DRAMP01423       | GLLSGILGAGKHICGGLSLR                   | Synthetic  | 3.071               |
| DRAMP18605       | VNWWKILAKIKVVK                         | Synthetic  | 3.597               |
| DRAMP18607       | VNWWKILPKIKVKV                         | Synthetic  | 4.332               |
| DRAMP03983       | KWWKLFKIPKFLHLA                        | Synthetic  | 4.332               |
| DRAMP03862       | RLRKIVVRVR                              | Frog       | 6.605               |
| DRAMP01310       | FLGLPSVSGAVSLKVL                      | Synthetic  | 7.090               |
| DRAMP04187       | KWLKILLKLL                              | Synthetic  | 8.957               |
| DRAMP03974       | WKKIPKFLHLAKKF                         | Fish       | 9.853               |
| DRAMP18601       | VNWWKILAKIKVAK                         | Synthetic  | 11.312              |
| DRAMP18602       | VNWWKILKKIKVAK                         | Synthetic  | 11.854              |
| DRAMP18603       | VNWWKILPKIKVAK                         | Synthetic  | 13.681              |
| DRAMP03981       | KWLKILPKFLHLAK                         | Synthetic  | 20.981              |
| DRAMP03850       | FALALKALKLKKKLLKKKKK                   | Scorpion   | 22.396              |
| DRAMP04002       | ILKILKGKFKK                            | Xenopus mueller | 22.471         |
| DRAMP18560       | JWRIFRRIRFRIF                          | Synthetic  | 22.608              |
| DRAMP03859       | RLARIVVIRWAR                           | Synthetic  | 22.632              |
| DRAMP03880       | RRIRWRIR                               | Synthetic  | 22.668              |
| DRAMP04318       | IKWKKLLRAAKRIL                         | Synthetic  | 22.671              |
| DRAMP18642       | KWRRIW                                 | Bacteria   | 22.674              |
| DRAMP04297       | LKALKLLAKKLKLL                         | Synthetic  | 22.677              |
| DRAMP04215       | RRLFRRILKWL                            | Synthetic  | 22.677              |
| DRAMP18514       | KKLHALHKLKWLHKLAKKKK                   | Synthetic-De Novo | 22.679        |
| DRAMP04384       | KIASIGKEIVKAL                          | Synthetic  | 22.679              |
| DRAMP03702       | LGGLPSLIGGLVSFK                        | Synthetic  | 22.679              |
| DRAMP02910       | LRRIRKIIHIKK                           | Synthetic  | 22.679              |
| DRAMP02911       | IRRIRKIIHIKK                           | Synthetic  | 22.725              |
| DRAMP01621       | IGPVLGVSGALGGGLKKI                     | Synthetic  | 22.837              |
| DRAMP01626       | IGPVLGVSGALGGGLKKI                     | Synthetic  | 24.947              |
| DRAMP04125       | SGKLWRRKKK                             | Frog       | 25.316              |
| DRAMP03975       | KFWKKIPKFLHLKKKF                       | Bombina variegata | 25.316        |
| DRAMP03973       | WFKKIPKFLHLAKKF                        | Synthetic  | 25.568              |
| DRAMP03979       | KWWKFLKFIPKFLHLAKKF                    | Orange-legged leaf frog | 25.616        |
| DRAMP01807       | FLPLVLGALGGIPKIL                       | Synthetic  | 29.105              |
| DRAMP03977       | WKKIPKFLHLKKKF                         | Synthetic  | 29.774              |
| DRAMP03976       | WFKKIPKFLHLKKKF                        | Synthetic  | 36.868              |
| DRAMP18619       | RKLRLKRIIAHKVKKKY                      | Synthetic  | 60.154              |
### Table S7. Newly Predicted Antibiofilm Peptides with MBIC Range 8–16 (µM) from the DRAMP Database

| Name        | Seq                        | Source          | Predicted MBEC (µM) |
|-------------|----------------------------|-----------------|---------------------|
| DRAMP02474  | KFFRKLKKSVKKRAK             | Spider          | 2.397               |
| DRAMP18600  | NVWKILGKIJKVK               | Synthetic       | 2.462               |
| DRAMP18568  | VSWKKSGLGKIJKVK             | Synthetic       | 3.071               |
| DRAMP18596  | NVWKVLGKIJKVAK              | Synthetic       | 3.071               |
| DRAMP04228  | GRFRRLGRFKKLFKKYGP          | Synthetic       | 3.071               |
| DRAMP03972  | KWFKIPKFLHLKK                 | Scorpion        | 3.236               |
| DRAMP02963  | GRFRRLKTRKRLKKIGKV          | Synthetic       | 4.011               |
| DRAMP01126  | ILGPVISKGVLGGKKNL            | Synthetic       | 4.332               |
| DRAMP03570  | LLGDFRKSKEKIGFKRIVQR        | Bovine          | 4.719               |
| DRAMP03967  | KWLFKKIPKFLHLKK             | Synthetic       | 4.953               |
| DRAMP03969  | KWLFKKISKFHLHAKK             | Synthetic       | 4.953               |
| DRAMP03970  | WLFLFKIPKFLHLHAKK             | Synthetic      | 6.610               |
| DRAMP03971  | FKLFKIPKFLHLHAKK             | Synthetic       | 10.699              |
| DRAMP18563  | KSLRRWVRSWR                  | Synthetic       | 10.791              |
| DRAMP03826  | KNLRRIHKRIHIKKY               | Odorana grahami (Frog) | 15.718          |
| DRAMP04310  | ELAKALKALKKALKSAR            | Synthetic       | 16.091              |
| DRAMP04311  | ELAKKALRAKALKSKASK           | Pig             | 19.720              |
| DRAMP03227  | GLFGKLIKFGKRAISKAVVKARGKH    | Synthetic       | 19.996              |
| DRAMP03980  | KWLFLFKKIPHLHAKF             | Toad            | 21.695              |
| DRAMP03923  | KWLFKKIGAVLKVLT              | Synthetic       | 22.608              |
| DRAMP04260  | LALLKVLRKIKKAK                | Synthetic       | 22.645              |
| DRAMP18562  | KSLRWRVSRW                   | Synthetic       | 22.645              |
| DRAMP18597  | VNWKVLAKIKVAK                | Synthetic       | 22.659              |
| DRAMP18598  | VNWKVLKIIJKVAK               | Synthetic       | 22.659              |
| DRAMP18599  | VNWKVLPKIIJKVAK              | Synthetic       | 22.668              |
| DRAMP03963  | KWLFLKKIKHLHSKAF              | Synthetic      | 22.671              |
| DRAMP03968  | KWLFLKKIKFLHLHAKF             | Synthetic       | 22.676              |
| DRAMP03984  | KWLFLKKIPHLKAF               | Synthetic       | 22.676              |
| DRAMP18570  | VNRRKILGSIKVVK               | Synthetic       | 22.679              |
| DRAMP18502  | WKSYYWRWR                   | Synthetic        | 22.679              |
| DRAMP04018  | ALYKFKKLLLKLKSLRKL          | Bacteria        | 22.679              |
| DRAMP04003  | ILGKIIKIKKL                 | Synthetic        | 22.679              |
| DRAMP04376  | JNWKLGIKKISAL                | Synthetic       | 22.705              |
| DRAMP18507  | FSGGNCRFRRRCFCTK             | Synthetic       | 22.725              |
| DRAMP04614  | AVNPEKHI.FRCKSICF           | Synthetic       | 22.837              |
| DRAMP03865  | RWKIVWRWR                   | Synthetic        | 22.837              |
| DRAMP18567  | VNWKILGSIKVK                | Synthetic        | 23.595              |
Table S8. Newly Predicted Antibiofilm Peptides with MBIC Range 8–16 µM from the DRAMP Database (Cont.)

| Name      | Seq                 | Source       | Predicted MBEC (µM) |
|-----------|---------------------|--------------|---------------------|
| DRAMP04664 | AKRLKKLAKKIWKWK     | Human        | 24.958              |
| DRAMP04051 | RICRIVVIRCIR        | Human        | 25.040              |
| DRAMP03978 | KWKLFKKIPFLHLAKKF   | Synthetic    | 25.219              |
| DRAMP03925 | KWKLFKKGAVLKVLT     | Synthetic    | 25.595              |
| DRAMP03768 | ILSAIWSGIKS         | Synthetic    | 25.595              |
| DRAMP03858 | RLARIVKIRVAR        | Synthetic    | 25.595              |
| DRAMP03982 | KWKLFKKIPHLAKKF     | Synthetic    | 28.344              |
| DRAMP18564 | KWLRRVWRWWWR        | Snake        | 35.077              |
| DRAMP18565 | KRLRRVWRWWWR        | Synthetic    | 39.599              |
| DRAMP03964 | KWKLFKKIPKFLHSAKKF  | Synthetic    | 39.927              |
| DRAMP03924 | KWKLFKKGIGAVLKVL    | Synthetic    | 44.372              |
| DRAMP03755 | FKRIVQRIKDFLIR      | AnTomato     | 48.038              |
| DRAMP02872 | GRFKFRKFKFLFKKLS     | Synthetic    | 50.828              |
| DRAMP18501 | WKSYSVRRWRS         | Synthetic    | 53.153              |
| DRAMP03829 | GLKKLKGKLKLGKLKLLL  | Synthetic    | 54.066              |
| DRAMP18499 | WKSYSVRRWRSR        | Synthetic    | 63.560              |

Table S9. Newly Predicted Antibiofilm Peptides with MBIC Range 16–32 (µM) from the DRAMP Database

| Name      | Seq                 | Source                      | Predicted MBEC (µM) |
|-----------|---------------------|-----------------------------|---------------------|
| DRAMP03777 | JWSAIWSGIKGLL       | *Unradicus yaschenkoi* (scorpion) | 35.514              |
| DRAMP04390 | INWKKGKEVLKAL       | Synthetic                   | 22.679              |
| DRAMP03966 | KLKLFKKIGIGKFLHSACKF | Synthetic                  | 7.220               |
| DRAMP02914 | RICRIIFLRVCR        | Sheep                       | 14.287              |
| DRAMP04377 | JWNLKGLKLKLSAL      | Synthetic                   | 22.679              |
| DRAMP04113 | KWKSFIKKLASKFLHSACKF | Synthetic                  | 22.680              |
| DRAMP04115 | KWKSFIKKLTKFLHSACKF  | Synthetic                  | 45.279              |
| DRAMP04127 | RGKRWWRRKK          | Synthetic                   | 49.900              |
| DRAMP04054 | RLCRIVVVIRVCR       | Synthetic                   | 47.017              |
| DRAMP18401 | ILSAIWSGIKGLL       | Scorpion                    | 35.514              |
| DRAMP03965 | KAKLFFKIGIGKFLHSACKF | Synthetic                  | 5.068               |
| DRAMP04102 | KWKSFIKKLTSKFLHLAKKF | Synthetic                  | 25.740              |
| DRAMP04103 | KWKSFIKKLTSKFLHSACKF | Synthetic                  | 45.279              |
| DRAMP04104 | KFKSFKKLTSKFLHSACKF  | Synthetic                  | 45.279              |
| DRAMP04106 | KWKSFIKKLTSKFLHSACKF | Synthetic                  | 45.279              |
| DRAMP04119 | KWKSFIKKLTSKFLHSKKKF | Synthetic                  | 45.279              |
| DRAMP04112 | KWKSFIKKLLSKFLHSACKF | Synthetic                  | 22.380              |
| DRAMP18723 | KWKLFKKI             | Moth                        | 48.126              |
| DRAMP18566 | VNWKSSLGKSIVVK      | Synthetic                   | 4.953               |
| DRAMP01134 | ILGPVIKTIGGVGLGILLKNL | Toad                      | 19.866              |
### Table S10. Newly Predicted Antibiofilm Peptides with MBIC Range >32 (µM) from the DRAMP Database

| Name     | Seq                        | Source     | Predicted MBEC (µM) |
|----------|----------------------------|------------|---------------------|
| DRAMP04343 | IGKLFKRIVKKILKFLRKL        | Synthetic  | 35.719              |
| DRAMP04126 | SGKRWRWRK                 | Synthetic  | 49.900              |
| DRAMP04339 | IGKKFKRIVQRIKKFLRKL       | Synthetic  | 3.556               |
| DRAMP04340 | IGKKFKRIVKRIKKFLRKL       | Synthetic  | 3.556               |
| DRAMP04345 | IGKWKRIJKKRIKKFLRKL       | Synthetic  | 3.556               |
| DRAMP04346 | IGKKFKRIVKRIKKWKLKL       | Synthetic  | 3.556               |
| DRAMP18614 | VRRFAWWAFLRR             | Synthetic  | 22.940              |
| DRAMP03876 | RRWWWRWRWRW              | Synthetic  | 49.900              |
| DRAMP03877 | KKWWWKWWK              | Synthetic  | 49.900              |
| DRAMP03878 | RRRWRRWRWRW             | Synthetic  | 49.900              |
| DRAMP03879 | RRRFRRFRF               | Synthetic  | 49.900              |
| DRAMP18457 | KRWWKWWRRRC            | Synthetic  | 15.828              |
| DRAMP18508 | KFAKKFKFFAAFKKFKK          | Synthetic  | 49.900              |
| DRAMP18643 | KWWWWRW                  | Synthetic  | 49.900              |
| DRAMP04001 | ILGKIWKIKSLF             | Synthetic  | 7.590               |
| DRAMP04069 | CFPFIFGSGFKFKFKFC         | Synthetic  | 22.791              |
| DRAMP04070 | CWWWRWGWGKWWK            | Synthetic  | 22.791              |
| DRAMP03869 | RRRWVRWRW              | Synthetic  | 40.529              |
| DRAMP18558 | FIKRIARLLRKIF           | Synthetic  | 34.455              |

### Table S11. Newly Predicted Antibiofilm Peptides with MBIC Range 1–8 (µM) from the UniProt database

| Name     | Seq                        | Source     | Predicted MBEC (µM) |
|----------|----------------------------|------------|---------------------|
| sp—P0CF03 | FLOGGILGPLMSLPGGLK        | Ant        | 19.723              |
| tr—A0A5K1B3V0 | VIRIGCKWKRTA      | Nymphaea colorata (plant) | 6.260          |
| tr—A0A0E9S200 | MCTRWRVLTCVRRR       | Anguilla anguilla (eel) | 28.711         |
| tr—A0A2P2N8A3 | MGKDFRFRFCHVKKVL   | Rhizophora macromonata (plant) | 11.138        |
| tr—A0A0A9M1Q7 | MGKKFKKWLKT        | Arundo donax (plant) | 14.310         |
| sp—C0HK43 | VNWKKILGKAKV       | Lasioglossum laticeps (bee) | 3.071         |
| tr—A0A0A9U210 | MTRRRRRHLLLRL | Arundo donax (plant) | 22.612         |
| sp—P17236 | FLPLLGLKVKGLLL      | Oriental hornet | 36.241          |
| tr—A0A2P2QY8 | MLKLRJLRLRKL         | Rhizophora macromonata (plant) | 35.601       |
| tr—A0A5K1FAL9 | FRABLLRTAFR         | Nymphaea colorata (plant) | 22.722         |
| sp—P82419 | GLOVDLGVGGLKGLLLP      | Ant        | 28.502              |
| sp—C0HL3D5 | FLSPKRJGAGASLWKL | Frog        | 23.319              |
| sp—P82420 | GLVDLGVGGLKGLLLPG      | Ant        | 27.711              |
| tr—E4Z311 | IKGILLRIKJKVR      | Oikopleura dioica (tunicate) | 35.514        |
### Table S12. Newly Predicted Antibiofilm Peptides with MBIC Range 8–16 (µM) from the UniProt database

| Name       | Seq                        | Source                        | Predicted MBEC (µM) |
|------------|-----------------------------|-------------------------------|---------------------|
| tr—A0A3D5SU75— | PCPCGSGKYYKHCHGKLS         | Rhodocyclaceae bacterium      | 1.854               |
| sp—P0C424—   | CCAPSCACRLGCRRCCR          | Conus marmoreus (marble cone) | 2.968               |
| tr—A0A5K1BN05— | LGCGLHPGIFACLK            | Nymphaea colorata (plant)     | 3.071               |
| tr—A0A5K1DCQ4— | LGCGLHPGIYACLK            | Nymphaea colorata (plant)     | 3.071               |
| sp—C0H4KI2—   | VNWKVYLGKJIVKAK            | Lastigosoma laticeps (bee)    | 3.071               |
| sp—P3O259—   | GCKRKRARKRPKKCRKARPKRKKVAKKCC | Catshark                      | 4.290               |
| sp—P14215—   | RRWCFRVCYRFYRCYRCR         | Atlantic horseshoe crab       | 6.030               |
| sp—P14216—   | RRWCFRVCYRFYRCYRCR         | Atlantic horseshoe crab       | 6.030               |
| tr—A0A033VX2— | AVPSWRIKSWNNR              | Bacteria                      | 7.090               |
| tr—E918P2—   | MLKKKLGSRSRRRCCRRRRRRRRA   | Solenopsis invicta (ant)      | 8.156               |
| sp—P69135—   | KWCFFRVCYRGICYRRCR         | Asian horseshoe crab          | 18.171              |
| sp—Q8WMD3—   | MARYRHRCSSRSRRCRRRRRRRCCRRRRRRRRA CCRYRCRRR | Bat                          | 21.567              |
| tr—A0A033H27— | MFGGGPLKLL                 | Oryza barthii (plant)         | 22.009              |
| tr—A0A1CDY26— | FLGCGRVQLAIKSIGI           | Triticum aestivum x Aeeglosa sharonensis (plant) | 22.359 |
| tr—A0A3Q7GQQ6— | GLAYRLYVNLHPCTKR           | Solanum lycopersicum (tomato) | 22.657              |
| tr—A0A01K7640— | KAIALAKGKCK                | Lycium cestroides (plants)    | 22.679              |
| sp—P0C20—    | INLKKIAKGLKSL              | Wasp                          | 22.679              |
| sp—P5878—    | INWKKIAISIGKEVLKAL         | Wasp                          | 22.679              |
| sp—P69034—   | INWLKLGKKVSAIL             | Wasp                          | 22.679              |
| sp—P69036—   | INWLKLGKAVDIAL             | Wasp                          | 22.679              |
| sp—P65443—   | GLDLDFLKAAGKLVRTNL         | Frog                           | 22.691              |
| tr—A0A5K0UX57— | ALLKSKPKLLRSGL             | Nymphaea colorata (plant)     | 22.694              |
| tr—E9جار4—   | KLVRRIILACLHIAVCK          | Solenopsis invicta (ant)      | 22.924              |
| sp—P82282—   | IGPVLGMVSAALGLKKIG         | Toad                           | 23.278              |
| sp—P82285—   | IGPVLGLVSAALGLKKIG         | Toad                           | 23.278              |
| sp—P85982—   | IFGAILPLALGALKNLIK         | Frog                           | 23.712              |
| LFB0040      | FKCRRWARMKMKLGA            | Synthetic                      | 25.902              |
| sp—P0C1R0—   | ILGTILGLKSL                | Wasp                          | 32.702              |
| tr—A0A5K0UVL7— | ILLKIVGCKIK                | Nymphaea colorata (plant)     | 38.404              |
| LFB0017      | FKCRRWQRWQVR               | Lactoferrin, source: cow      | 38.492              |
| sp—C0HLM2—   | SGCCCHPACGKKNRC            | Alpha-conotoxin, source: conus purpurascens | 39.130 |
| tr—A0A0K0L6U6— | GOCYCCGAFQRQICYRKY         | Androctonus biclor (scorpion) | 46.616              |
| tr—A0A5K1IP988— | EFKIHKSGRRKWM           | Nymphaea colorata (plant)     | 46.812              |
| tr—I7DAN2—   | KSKGKIKSKAYKR              | Rhea americana (bird)          | 50.828              |
| tr—S7JKV4—   | SLFCKGCSKL                 | Chlamydia psittaci (bacteria)  | 60.154              |
| tr—Q16228—   | WRWRACRRIPGRPFWRV          | Human                          | 61.163              |

### Table S13. Newly Predicted Antibiofilm Peptides with MBIC Range 16–32 (µM) from the UniProt database

| Name       | Seq                        | Source                        | Predicted MBEC (µM) |
|------------|-----------------------------|-------------------------------|---------------------|
| sp—P0C05—   | GLLKRIKTL                  | Wasp                          | 22.679              |
| sp—P17238—  | INLKIAAALVKKVL              | Hornet                         | 22.679              |
| tr—Q9USM9—  | AGLGGICGLDTNREIVKSGPK       | Scaptomyza granimum (insect)   | 20.318              |
## 7 DATASET

### 7.1 Positive Dataset

The details of our positive dataset, including the peptide sequence and its length, are given in Tables S14–S18.

| Name                        | Seq                                           | Seq Length |
|-----------------------------|-----------------------------------------------|------------|
| BREVININ-1GHA               | FLGAVLKVGKLVPAAICKISKKC                       | 24         |
| DERMASEPTIN-AC4             | SLWGGKLEMAAAGKAALNAVGLNQ                      | 27         |
| RPDEF1ALPHA                 | GFGCPNDYSCSNHCRCDSIOCRRGGYCKYHVICTCYGCKKRSIQE | 44         |
| KASSINIATUERIN-3            | FQHLILPHAQIQIKDIF                             | 20         |
| AGELAIA-ADN                 | IWLKIGKAIDAL                                  | 14         |
| CCL20                       | SNFDCCLGTYDRLHPKFIVGTRQLANEGCDINAIIFHTKKKLSCVANCANPQTWVKYIVRRNSSKVKNM | 69         |
| CHICKEN                     | RFGFLKLRIRRPKFYVTITIQGSARFG                   | 27         |
| CITROPIN                    | GLFDVKVAVSGGL                                  | 16         |
| COLISTIN                    | KTJAAAALKK                                    | 10         |
| CON10                       | FWSFLYKAAASKKLPSILIGDNDKSSS                   | 27         |
| COPRISIN                    | VTCDFYLSEAKVIAYNHSAACALRLRKKKGSCQNCVCRN       | 43         |
| DATUCIN                     | TFPKCAPTRPGPKCDINNFKSFHIFIWRA                 | 31         |
| DERMASEPTIN-PH              | ALWKEVLNKAGAKALNENLVL                         | 22         |
| DERMASEPTIN-PT9             | GLWSIKDAAGTAGKAALGFVNEMV                      | 25         |
| DHVAR4                      | RKLFFKLLSFLKRKY                                | 14         |
| ENTEROCIN                   | LGSVANKIKDIPEFAMISIAVKAQQKAWKELAVTVMFAKANGLKTNAIVAGQLAWQCQLS | 68         |
| ESCULENTIN                  | GFSKLAGKIKLLNSILKGL                           | 21         |
| GL13K                       | GJIJKLAKSLKL                                  | 13         |
| GRAMICIDIN                  | VKLFPVKLF                                     | 10         |
| HS02                        | KWAVRIRKHFKGFIS                               | 16         |
| HUMAN defensin              | GIINLTKCYVRGRCGVLSCLKPEEKIQICKCSTRGRKCCRKK    | 45         |
| HYICIN                      | NKCNGCACAIAACADGPIDFVEVGATITFGGI             | 35         |
| INDOLICIDIN                 | ILPKWDPWVWPWRR                                | 13         |
| JAPONICIN-2LF               | FIVPSIFLLKAFIALKKC                            | 20         |
| LL-37                       | LLGDFFRSKEKIGKEKFRIVQIRKDLRLNLPRTES           | 37         |
| MP-C                        | LNLKALLAVAKKL                                 | 14         |
| MORONECIDIN-                | FFRNLWKGAKAARAFRAHAAWRA                       | 22         |
| MYXINIDIN                   | GHIILKYGKPS                                   | 12         |
| NA-CATH                     | GLLSGILGAGKIKVF                               | 15         |
| NISIN                       | ITSISLCTPOCKTGALMGCNMTATCHCSIHVSK             | 34         |
| PARACENTRIN                 | EVAADFSLKLL                                  | 11         |
| PHYLLOSEPTIN-1              | FLSHIPHVGSVIAKHF                              | 19         |
| PHYLLOSEPTIN-1-CO           | FLSMPKJAGGIASVLKNL                            | 19         |
| PHYLLOSEPTIN-PHA            | FSLIPAAISAVSALANHF                            | 19         |
| PLEUROCIDIN                 | GWGFSFKKAHVHGKVKHGAALTHYL                    | 25         |
| POLYBIA-MP-II               | IWLNLGKMVIDAL                                 | 14         |
| POLYMYXIN                   | KTKKKFLKKT                                   | 10         |
| PROTEGRIN                   | RGRRLCHYRRCIFCIVCR                           | 18         |
| SA-CATH                     | KFFKLKKSVKKHVVKKFFKPKVIGVSIHF                | 30         |
| SAAP-148                    | LKKRVKRRFLKLRKRYRQLKKPVRR                    | 24         |
| SMAP-29-APD                 | RGLRLRGLKIAHGYKIVQPTVLRIRIAG                | 29         |
| TACHYPLESIN                 | KWCFCRVCYRPIRCYRKCR                          | 17         |
| TEMPORIN-1OLA               | FLPLLXILGKIL                                 | 13         |
| Name                | Seq                        | Seq Length |
|---------------------|----------------------------|------------|
| TEMPORIN-B          | LLPIVGNLLKSSL             | 13         |
| TEMPORIN-1CEB       | ILPILIGGLGGLBK            | 14         |
| TEMPORIN-GHC        | FLQHIIGALTHIF             | 13         |
| TEMPORIN-GHD        | FLQHIIGALSHFF             | 13         |
| TEMPORIN-PTE        | FFGSVKLKIPKIL             | 13         |
| TETRAF2W-RK         | WWWLRIKIV                 | 8          |
| TOAP1               | FIGMIPGLIGGLISAK          | 17         |
| TOAP2               | FFGTLKGLSLKLPGK           | 26         |
| TSAP-2              | FIGMIPGLIGGLISAFK         | 17         |
| UYCT3               | ILSAIWSGIKSLF             |            |
| VLL-28              | VLVVTLTRLHQRGVIYRKRHRFSGRKYR | 28         |
| ZMD32               | RTCQGSQHREFGIPCLRSNCANVCRTEGFPGGRGCFRRRCCTTCRC | 47         |
| BMAP-27             | GRFKRKKKFFKLLLKSPVIIPLLHL | 26         |
| BMAP-28             | GGLRLGRKILRAWKKGVPPIRIR   | 27         |
| SMAP-29             | RGLLRGLIRKIAHVKKYGPTVLRIRIA | 28         |
| KSL                 | KKVFFKVKF                 | 10         |
| F2-5-12W            | RWGRWLKIRRRWPRK           | 15         |
| LL-31               | LLGDFFRKSSKEKIGKEFKRIVQRKDFRLNL | 31         |
| LL-31               | RKSKEKIGKEFKRIVQRKDFRLNL  | 25         |
| LL-37               | IKGKEFKRIVQRKDFRLNLVPRTES | 25         |
| LL-37               | RKSKEKIGKEFKRIVQRKDFRLNLVPRTES | 31         |
| LL-19               | LLGDFFRKSSKEKIGKEFKR      | 19         |
| LL-25               | LLGDFFRKSSKEKIGKEFKRIVQIK | 25         |
| LL-13               | LLGDFFRKSSKEK             | 13         |
| LL-13-31            | IKGKEFKRIVQRKDFRLNL       | 19         |
| LL-7-25             | RKSKEKIGKEFKRIVQRQIK      | 19         |
| LL-13-25            | IKGKEFKRIVQRQIK           | 13         |
| LL-19-37            | RIVQRIKDFRLNLVPRTES       | 19         |
| KSL-W               | KKVFFWVKF                 | 10         |
| KS-30               | KSKEKIGKEFKRIVQRKDFRLNLVPRTES | 30         |
| KR-20               | KRIVQRKDFRLNLVPRTES       | 20         |
| KR-12               | KRIVQRKDFRL              | 12         |
| LACTOFERRICIN-(17-30) | FKKRRWQWRMMKKLG      | 14         |
| LACTOFERRAMPIN      | WKLISQAQKFGKNSR           | 17         |
| MUC7-12-MER-L       | RKSYYKLHKRCR              | 12         |
| G10KHC              | KKHRRKHRRKHGSGGSKNLRRIRKGIHIKKYG | 36         |
| MUC7-12-MER-L4      | RKSYYKHALKRAR             | 12         |
| MUC7-20-MER         | LAHKQPFIRSKSYKCLHKRCR     | 20         |
| HSN5                | AKRRHGGYKRFH              | 12         |
| MAGAININ-II         | GIGKFLHSAKFKGAFVGEIMNS    | 23         |
| LYS-A1              | KIFGAIWPLGALKNLKJ        | 19         |
| AAP2                | FHFFHFFHFHFFH           | 14         |
| CSP                 | SGSSLSTFFRFLGNGSFTQALGK   | 21         |
| CSPC16              | TFFRFNSRFFTQLGK          | 16         |
| G2                  | KNLRIRKGIHHK              | 16         |
| C16G2               | TFFRFNSRFFTQLGKGGGBKNLRIKIRKGIHIKKY | 35         |
| M8G2                | TFFRFNSRGGKRLRIRKGIHIKKY | 27         |
| S6L3-33             | FKKFVKWRFRF              | 11         |
| C16-33              | TRRRLFNSRFFTQALGKGGGGFFKKFWKFWRF | 31         |
| M8-33               | TFFRFNSRGGGFFKKFWKFWRF   | 23         |
| CECROPIN-A-(1-7)-MELITTIN | KWKLFFKIGAVLKL   | 15         |
| HH15                | KRRFRIVRVRK              | 12         |
### Table S16. Peptide List for our Positive Dataset (Cont.)

| Name          | Seq                              | Seq Length |
|---------------|----------------------------------|------------|
| BAC2A         | RLARIVVIRVAR                     | 12         |
| 1026          | VQWRIRVRIHK                      | 12         |
| 1029          | KQFRIRVRV                        | 9          |
| 1036          | VQFRIRVIRIK                      | 13         |
| 1037          | KRFIRVRV                         | 9          |
| HH2           | VQLRIRVAVIRA                     | 12         |
| 1002          | VQRWLVWRIRK                      | 12         |
| 1003          | IVWKIKRWVVGVR                     | 12         |
| 1004          | RFWKVVRVYIRF                     | 12         |
| 1008          | RIKWIVRFR                        | 9          |
| HH7           | VRLRIRVAVRA                      | 12         |
| 1010          | IWRIRVWVRRR                      | 12         |
| 1011          | RWWVVRIVQRR                      | 12         |
| 1012          | IFWRRIVIVKKF                     | 12         |
| 1013          | VRLRIRVA                         | 8          |
| 1016          | LRIRWFKR                         | 9          |
| HH8           | VRLRIRVAVIRK                     | 12         |
| 1020          | VRLRIRWWVLK                      | 12         |
| HH10          | KRFIRVAVRA                       | 12         |
| 1035          | KRWBVIVNIRR                      | 12         |
| 1031          | WRWWRVRWR                        | 9          |
| IMB-2         | TFFRLFNRGGGWGSGFFKKAAHVGVKL       | 25         |
| BAC8C         | RIWVTWR                         | 8          |
| PTP-7         | FLGALFKALSKLL                    | 13         |
| HOLOTHURIDIN-1| HLGHHALDHLKK                     | 12         |
| HOLOTHURIDIN-2| ASHGLGHALDHLKK                   | 14         |
| TN-AFP1       | LMCTHIRPLDCSN                    | 11         |
| COPRISIN-BAAMP| VTCDVLSFEAKGLAVNH                | 17         |
| HISTATIN5     | DSHAKRHRGKHFHKFHSHRGY            | 24         |
| HST           | AKRRHGYKKFHHGG                  | 15         |
| DF17-6K       | KKKKKAAFAAFAAFAA                 | 17         |
| DF21-10K      | KKKKKKKKKAAFAAFAAFAA             | 21         |
| CWR11         | CWFWKWKWRRRR                     | 12         |
| CHRYSOPTIN-1  | FFGWLIKGAIHAGKAIHGLIHRRRH       | 25         |
| RK1           | RWKRWWRRRRK                      | 10         |
| RK2           | RKKRWWRRRRK                      | 10         |
| (IRIK)        | IRIKIRIK                        | 8          |
| (IRVK)        | IRVKIRVKIRVK                    | 12         |
| ALPHA-DEFENSIN-3| DCYCRIPIACGERRYGTCTYQGRLWAFC   | 30         |
| BETA-DEFENSIN-1| DHYNCVSSGQCLYSACPIFTKIQGTCYRGKACCK | 36       |
| MAGAININ-1    | GIGKFLHSAGKFGKAFVGEIMKS         | 23         |
| RIP           | YSPWTNF                          | 7          |
| K4-S4(12-13)A| AWKTLKKVLKAA                     | 13         |
| DD13-RIP      | AWKTLKKVLKAYSPWTNF               | 20         |
| 2C-4          | RWRRWRF                          | 7          |
| SM6(L1)2C     | FIKHIHRFGGGRWRWRWF               | 19         |
| SM6(L3)2C     | FIKHIHRFSATRWRWRWF               | 19         |
| SM6(L1)B33    | FIKHIHRFGGGFKFKWKFWRFRF          | 23         |
| NRC-16        | GWKKWLRGBKAKHLGQAAIK             | 19         |
| GK7           | GQINLKK                         | 7          |
| (RW)2-NH2     | RWRRW                            | 4          |
| (RW)3-NH2     | RWRRWRW                          | 6          |
Table S17. Peptide List for our Positive Dataset (Cont.)

| Name                      | Seq                | Seq Length |
|---------------------------|--------------------|------------|
| (RW)4-NH2                 | RWRWRWRWRW         | 8          |
| LASIO-III                 | VNWKKILGKIIKVVK    | 15         |
| MELITLIN                  | GIGAVLKVLTTLPAIJSWKVRKQQ | 26         |
| MELIMEINE                 | TLISWIKNRKQRPVRVRRRRRRRGGRRRRR | 29         |
| MELIMEINE-CYNS            | CTLISWIKNRKQRPVRVRRRRRRRGGRRRRR | 30         |
| MELIMEINE-CYS             | TLISWIKNRKQRPVRVRRRRRRRGGRRRRC | 30         |
| MELIMEINE-CYS13           | TLISWIKNRKQCRPVRRRRRRRGGRRRRR | 30         |
| K4-S4(1-15A)              | LWKTLKKVVLKAAA     | 14         |
| BETAA6-20-G3K6            | NEEGFSARGHRPLDGGGKKKKK | 24         |
| HEPCIDIN                  | ICIFCGCCSHSRKCGMCCT | 20         |
| NA-CATH-BAAMP             | KRFFKKFFKLLKNSVKKRACKFKKPKVIGVTFFP | 34         |
| NA-CATH-ATRA1-ATRA1       | KRFFKKFFKLLKNSVKKRFFKKFKLKVIGVTFFP | 34         |
| LACTOFERRICIN-B-(17-41)   | FKCRWRQWRMKKLGAPSITCVRAF | 25         |
| SCRAMBLED                 | GLKLRFEFSKIKGFLKTEPVFRDFDIKLKDNRISVQR | 37         |
| R-FV-16                   | RFRLFRIRVRLKKI     | 16         |
| FV7                       | FRIRVRV           | 7          |
| VSL2                      | AFKAFWKVFVFKVFK    | 13         |
| VS2                       | KWFWKVFVFKVFK      | 11         |
| L-K6                      | IJKILSKIKLLK       | 13         |
| HLF1-11                   | GRRRSSVQWCA        | 11         |
| FS3                       | YAPWNTF           | 7          |
| TET-213                   | KRWWKWWRRC        | 10         |
| 1010CY5                   | IWRIRVRVWRRRC     | 13         |
| TET-20                    | KRWRIRVRVIRKC     | 13         |
| TET-26                    | WIVVIWRRRKRRRC    | 13         |
| FS8                       | YAPWNTA           | 7          |
| CHROMOFUNGIN              | RILSILRHQNLKELKDLAL | 20         |
| CECROPIN-B                | KWKVFKKIEKMGRNIRNGIVKAGPAAVLGEAKAL | 35         |
| MAGAININ                  | GIGFLHSAGLFGLFVGEIMKS | 23         |
| CYSLASIO-III              | CVNWKKILGKIIKVVK  | 16         |
| DASAMP1                   | FFKVLKLRKIF       | 13         |
| BMAP-18                   | GRWRKRWKKWKLWKQLS | 18         |
| BACTENECIN                | LRCRIVVIRVCR      | 12         |
| CA-MA                     | KWKLFKKIGIGKFHLSAKKF | 20         |
| RTA3                      | RPAKRKAARVRMACV   | 16         |
| DHVAR5                    | LLLLFLKKRKKRKY     | 14         |
| KABT-AMP                  | GIWKKWIKWLKWKWLKLWKKG | 22         |
| P10                       | LAREYKIKVEKLKRWLRQVLRLTR | 24         |
| P60.4AC                   | IGKFKRIVERIKRFLRELVRPLR | 24         |
| OSIP108                   | MLCVLQGLRE        | 10         |
| S-OSIP108                 | ELRLVCMGQL        | 10         |
| [CYC2]OSIP108             | MLCVLQGLREGG      | 12         |
| [CYC3]OSIP108             | MLCVLQGLREC       | 11         |
| I018                      | VRLIHAVRIWRR      | 12         |
| HE1                       | RRWRVAVILRV       | 12         |
| HE2                       | VRLIARAVRAWV      | 12         |
| HE3                       | VRWARVARILRV      | 12         |
| HE4                       | VRLIHAVRIWRR      | 12         |
| HE10                      | VRLIHAVRIWRR      | 10         |
| HE12                      | RFKRVARFWV        | 10         |
| GL13KR1                   | IGKLLKSKLKLAL     | 13         |
| [IKIK]2                   | IKIKIK             | 8          |
### Table S18. Peptide List for our Positive Dataset (Cont.)

| Name         | Seq                      | Seq Length |
|--------------|--------------------------|------------|
| RI1012       | FKKVIVIRRWF1             | 12         |
| RI1002       | KRIRWVILVRQV             | 12         |
| LJK1         | VFLRIRIVIVIR             | 12         |
| RIJK1        | RIVIVRIRRLFV             | 12         |
| LJK2         | VFWRIRVWVIR              | 12         |
| RIJK2        | RIVVWRIRRWVF             | 12         |
| LJK3         | VQLRAIRVVRV              | 12         |
| RIJK3        | RIVVRIARLQV              | 12         |
| LJK4         | VQLRIRVWVIR              | 12         |
| RIJK4        | RIVVWRIRRLQV             | 12         |
| LJK5         | VQWRIRVVRV               | 12         |
| RIJK5        | RIVVRAIRWQV              | 12         |
| LJK6         | VQWRIRVWVIR              | 12         |
| RIJK6        | RIVVWRIRWQV              | 12         |
| NAL-P-113    | AKRRRGYKRKFKKK            | 13         |
| P15          | GTGPGQIAGQRGVV            | 15         |
| P15-CSP      | GTGPGIAGQQRGVAEAAAEAAKAASGLSLTFRLFNRSFTQALGK | 53         |
| C-GG-NT-DHVAR5 | CGGLLLFLKKRRKRRKRY     | 17         |
| KT2          | NGVQPKYKWKKWKWW           | 17         |
| RT2          | NGVQPKYBWWRWWRW           | 17         |
| LF11-322     | FFWRIRIRR                | 9          |
| LF11-324     | FFWRIRIRR                | 10         |
| 6-MO-LF11-227 | FWRRFWR                    | 8          |
| LF11-215     | FWRRIRIRR                | 8          |
| D-ATRA-1A    | KRAKKFKKKLK              | 11         |
| ATRA-2       | KRAKKFKKKP               | 11         |
| ATRA-1       | KRFKKFKKKLK              | 11         |
| ALL          | LKKKLLLKKLKLKKL           | 15         |
| SEG5D        | KKKLLLLLLLLLKKK           | 15         |
| SEG6D        | LLLLLKKKKKKLLL            | 15         |
| G10          | KNLRRIRKGIHJIKKYG        | 18         |
| LIN-SB056    | WKKIRVRLSA               | 10         |
| LIN-SB056-1  | WKKIRVRLSA               | 10         |
| MYXINIDIN2   | KIKWILKYWKWS             | 12         |
| MYXINIDIN3   | RIRILRYYRSW              | 12         |
| GH12         | GLLWHLHLLLH              | 12         |
| PA-MAP       | LAAKLTKAAATKTLTAALTKLAAALT | 24         |
| HSAFP1       | DGYKLCDPVSPTSAGHCSSSKCSCQQCKDREHFAVGACPHQPSVCKCIRKQC | 54         |
| HSLIN06      | EHFAYGAKHYQPSVKKFKKRQK   | 24         |
| Verine       | RRRWWWV                  | 8          |
| Phylloseptin-PTa | FLSLIPKIAGGIAALAKHL    | 19         |
## 7.2 MBEC Dataset

Antibiofilm peptides with MBEC values are listed in Tables S19–S20. The pathogens against which the MBEC values are effective are also listed in the ‘pathogen’ column. The MBEC values are listed in µM.

### Table S19. Antibiofilm Peptides and MBEC (µM) Values

| Name                   | Seq                                             | MBEC (µM) | Pathogen   | Source             |
|------------------------|-------------------------------------------------|-----------|------------|--------------------|
| BREVININ-1GHA          | FLGAVLKQAVKLVPAAICKSKKC                        | 16        | S. aureus  | Chen et al. (2018) |
| DERMASEPTIN-AC4        | SLWGLKLEMAAAXAAGKANNAVNGLNQ                    | 256       | S. aureus  | Gong et al. (2020) |
| KASSIMIAUERIN-3        | FHLILPEIHPHQIKDRIF                             | 64        | S. aureus  | Wang et al. (2020) |
| CCL20                  | SNFDCCLGYTDRLHPKFQGFTQRLANEGDGINAFAHTTKKLSSCVANCPKQWGTVKYIVRLSSLKKVKNM | 128       | P. aeruginosa | Ramamourthy et al. (2019) |
| COPRISIN               | VTDVLSPAEKGAIVHNSACALHICARKKGGSCQNGVCCMN      | 4.49      | P. aeruginosa | Hwang et al. (2013) |
| DERMASEPTIN-PH         | ALWKEVLKNAGKAALEINNLV                          | 128       | S. aureus  | Huang et al. (2017) |
| DERMASEPTIN-PT9        | GLWSKIKDAAKT7AGKAAALGFPENMV                    | 32        | S. aureus  | Li et al. (2019)   |
| ESCULENTIN             | GIFSKLPAGKKINNLISLGK                           | 6         | P. aeruginosa | Luca et al. (2013) |
| OL13K                  | GKIHKLASKLKL                                     | 22.47     | P. aeruginosa | Hirt and Guo (2013) |
| HUMAN defense          | QINTLQKYVCVRGGRGCAVLSCLPKQQEQRGCSITGRKCRKK     | 3.1       | MRSA       | Sutton and Pritts (2014) |
| INDOLICIDIN            | ILPKWKKPPWPRK                                   | 335.7     | MRSA       | Mataraci and Dosler (2012) |
| LL-37                  | LLGDFRKSKEKIKFEKRIQKFKLKDRLNVRPSTES             | 20        | P. aeruginosa | Nagant et al. (2012) |
| NISIN                  | HTSISLCPKCIHTALMCMNIAHTCCSHYSK                 | 183.1     | MRSA       | Mataraci and Dosler (2012) |
| PHYLLOSEPTIN-1         | FLSLPHIVSGVASSAHLKF                           | 5         | S. aureus  | Zhang et al. (2010) |
| PLEUROCIDIN            | GWGPSFKKAHAVGKHAVGAALTHYL                     | 23.62     | S. mutans  | Tao et al. (2011)   |
| BMAP-27                | GFRKFKFKKFKLKLSPVPLHHL                         | 6.2       | P. aeruginosa | Pompiio et al. (2011) |
| BMAP-28                | GQLSRLGRKILRAWKKYGIPIVPHIR                    | 6.5       | P. aeruginosa | Pompiio et al. (2011) |
| SMAP-29                | RQLRLRGIKAHHVGYKQPVTLLRIIRIA                  | 6.25      | P. aeruginosa | Pompiio et al. (2011) |
| KSL                    | KKVVFVKFK                                       | 400       | S. mutans  | Liu et al. (2011)   |
| P2-5-12W               | RWGRWLKIRIRWRPK                                 | 40        | S. epidermidis | Mohloek et al. (2011) |
| LL-31                  | LLGDFRKSKEKIKFEKRIQKFKLKDRLNVRPSTES             | 20        | P. aeruginosa | Nagant et al. (2012) |
| LL-13-37               | 1GFEKRIQKFKLKDRLNVRPSTES                       | 100       | P. aeruginosa | Nagant et al. (2012) |
| LL7-37                 | RRSKEKIGEKFRIQKFKLKDRLNVRPSTES                 | 50        | P. aeruginosa | Nagant et al. (2012) |
| KSL-W                  | KKVVFVKFK                                       | 191.16    | MRSA       | Gwawde et al. (2014) |
| MUC7-12-MER-L          | RKSYKCLKHCR                                    | 12.5–50; median 50 | S. mutans | Wei et al. (2006) |
| MUC7-12-MER-L4         | RKSYKALHRAR                                     | 50; median >50 | S. mutans | Wei et al. (2006) |
| MUC7-20-MER            | LAHQKFIRKSYSKCLKHCR                             | 6.25–25; median 25 | S. mutans | Wei et al. (2006) |
| HSN5                   | AKRHHGYKRFH                                    | 12.5–50; median >50 | S. mutans | Wei et al. (2006) |
| MAGAININ-II            | GIGKFLHSKAKFGKAFVGEIMNS                        | 25–50; median >50 | S. mutans | Wei et al. (2006) |
| C16G2                  | TFFRLFNKSTFQALKGKGGKLNRIKRGHIHHKY               | 25        | S. mutans  | Sullivan et al. (2011) |
| BAC5C                  | RIWVIWR                                        | 108.13    | S. mutans  | Ding et al. (2014)  |
| PTP-7                  | FLGALFKALSKL                                   | 40        | S. aureus  | Kharidza and Liang (2011) |
| CHRYSOPHSIN-1          | FPGWLIKGAHAIKGAHIGLHIIHRK                     | 11.07     | S. aureus  | Wang et al. (2012)  |
| 2C-4                   | RWRWRWF                                        | 50        | S. mutans and other oral streptococci | He et al. (2010) |
| SM6s(L1)2C             | FIKHIHPRFQGGGRWRWRF                            | 50        | S. mutans and other oral streptococci | He et al. (2010) |
| SM6s(L3)2C             | FIKHIHPRFSATRWRWRF                             | 50        | S. mutans and other oral streptococci | He et al. (2010) |
| SM6s(L1)B3             | FIKHIHPRFQGGGRFKFKWWRF                        | 50        | S. mutans  | He et al. (2010)   |
| (RW)4-12H              | RWRWWRWF                                       | 100       | E. coli    | Hou et al. (2010)  |
| MELITIN                | GIGVAVLKLYVTGGDLPSIKKQRQQ                     | 50        | S. mutans  | Sullivan et al. (2011) |
| K-FV-116               | KFRKLFIRVRVVLKK                                | 64        | P. aeruginosa | Xu et al. (2014)   |
| L-K6                   | IKIKLSSIKKLL                                    | 6.25      | S. mutans  | Shang et al. (2014) |
| CA-MA                  | KWKLFFKIKGIGFKLHASKKF                          | 361.54    | MRSA       | Mataraci and Dosler (2012) |
| PI0                    | LAREYKKIVELKRRWLRQRVLRTL                       | 32        | HAASNA     | Haassna et al. (2014) |
Table S20. Antibiofilm Peptides and MBEC (µM) Values (Cont...)

| Name     | Seq              | MBEC (µM) | Pathogen      | Source                                      |
|----------|------------------|-----------|---------------|---------------------------------------------|
| 1018     | VRLIVAVRIWRR     | 6.51      | *P. aeruginosa* | de la Fuente-Núñez et al. (2013)             |
| HE4      | VRLIVAVRIWRR     | 6.16      | *P. aeruginosa* | de la Fuente-Núñez et al. (2014)             |
| HE10     | VRLIVRIWRR       | 7.32      | *P. aeruginosa* | de la Fuente-Núñez et al. (2014)             |
| DJK5     | VQWRARIRVIR      | 1.61      | *P. aeruginosa* | de la Fuente-Núñez et al. (2015)             |
| DJK6     | VQWRIRKRWVIR     | 1.5       | *P. aeruginosa* | de la Fuente-Núñez et al. (2015)             |
| KT2      | NVQIPRYKWWKWWKKWW| 1         | *E. coli*      | Anunthawan et al. (2015)                    |
| RT2      | NVQIPRYRWWRRRWW  | 1         | *E. coli*      | Anunthawan et al. (2015)                    |
| DL-MB-LF11-322 | PFWRIRIRR    | 246.38    | *P. aeruginosa* | Sánchez-Gómez et al. (2015)                 |
| LF11-324 | PFWRIRIRR        | 55.33     | *P. aeruginosa* | Sánchez-Gómez et al. (2015)                 |
| 6-MO-LF11-227 | FWRRFWRR    | 489.01    | *P. aeruginosa* | Sánchez-Gómez et al. (2015)                 |
| LF11-215 | FWRRIRRR        | 133.41    | *P. aeruginosa* | Sánchez-Gómez et al. (2015)                 |
| P60.4AC  | IGKEFKRIVERIKRFLRELVRPLR | 32 | MRSA | Haisma et al. (2014) |

REFERENCES

Anunthawan, T., de la Fuente-Núñez, C., Hancock, R. E., and Klaynongsruang, S. (2015). Cationic amphipathic peptides kt2 and rt2 are taken up into bacterial cells and kill planktonic and biofilm bacteria. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1848, 1352–1358. doi:https://doi.org/10.1016/j.bbamem.2015.02.021

[Dataset] Armstrong, D. and Zidovetzki, R. (2009). Helical wheel projections. www.donarmstrong.com/cgi-bin/wheel.pl. Version ID: wheel.pl,v 1.4 2009-10-20 21:23:36 don Exp

Chen, Q., Cheng, P., Ma, C., Xi, X., Wang, L., Zhou, M., et al. (2018). Evaluating the bioactivity of a novel broad-spectrum antimicrobial peptide brevinin-1gha from the frog skin secretion of hylarana guentheri and its analogues. *Toxins* 10. doi:10.3390/toxins10100413

de la Fuente-Núñez, C., Mansour, S. C., Wang, Z., Jiang, L., Breidenstein, E. B., Elliott, M., et al. (2014). Anti-biofilm and immunomodulatory activities of peptides that inhibit biofilms formed by pathogens isolated from cystic fibrosis patients. *Antibiotics* 3, 509–526. doi:10.3390/antibiotics3040509

de la Fuente-Núñez, C., Reffuveille, F., Mansour, S. C., Reckseidler-Zenteno, S. L., Hernández, D., Brackman, G., et al. (2015). D-enantiomeric peptides that eradicate wild-type and multidrug-resistant biofilms and protect against lethal *Pseudomonas aeruginosa* infections. *Chem Biol* 22, 196–205

de la Fuente-Núñez, C., Reffuveille, F., nde, L. F., and Hancock, R. E. (2013). Bacterial biofilm development as a multicellular adaptation: antibiotic resistance and new therapeutic strategies. *Current Opinion in Microbiology* 16, 580–589. doi:10.1016/j.mib.2013.06.013. Antimicrobials · Genomics

Ding, Y., Wang, W., Fan, M., Tong, Z., Kuang, R., Jiang, W., et al. (2014). Antimicrobial and anti-biofilm effect of Bac8c on major bacteria associated with dental caries and Streptococcus mutans biofilms. *Peptides* 52, 61–67

Gawande, P. V., Leung, K. P., and Madhyastha, S. (2014). Antibiofilm and antimicrobial efficacy of dispersinb®-ksl-w peptide-based wound gel against chronic wound infection associated bacteria. *Current Microbiology* 68, 635–641. doi:10.1007/s00284-014-0519-6

Gong, Z., Pei, X., Ren, S., Chen, X., Wang, L., Ma, C., et al. (2020). Identification and rational design of a novel antibacterial peptide dermaseptin-ac from the skin secretion of the red-eyed tree frog agalychnis caldidyras. *Antibiotics* 9. doi:10.3390/antibiotics9050243

Gupta, S., Sharma, A. K., Jaiswal, S. K., and Sharma, V. K. (2016). Prediction of biofilm inhibiting peptides: An in silico approach. *Frontiers in Microbiology* 7, 949. doi:10.3389/fmicb.2016.00949

Haisma, E. M., de Breij, A., Chan, H., van Dissel, J. T., Drijfhout, J. W., Hiemstra, P. S., et al. (2014). LI-37-derived peptides eradicate multidrug-resistant staphylococcus aureus from thermally wounded human skin equivalents. *Antimicrobial Agents and Chemotherapy* 58, 4411–4419. doi:10.1128/AAC.02554-14
He, J., Yarbrough, D. K., Kreth, J., Anderson, M. H., Shi, W., and Eckert, R. (2010). Systematic approach to optimizing specifically targeted antimicrobial peptides against *Streptococcus mutans*. *Antimicrobial Agents and Chemotherapy* 54, 2143–2151. doi:10.1128/AAC.01391-09

Hirt, H. and Gorr, S.-U. (2013). Antimicrobial peptide gl13k is effective in reducing biofilms of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy* 57, 4903–4910. doi:10.1128/AAC.00311-13

Hou, S., Liu, Z., Young, A. W., Mark, S. L., Kallenbach, N. R., and Ren, D. (2010). Effects of trp- and arg-containing antimicrobial-peptide structure on inhibition of *Escherichia coli* planktonic growth and biofilm formation. *Applied and Environmental Microbiology* 76, 1967–1974. doi:10.1128/AEM.02321-09

Huang, L., Chen, D., Wang, L., Lin, C., Ma, C., Xi, X., et al. (2017). Dermaseptin-ph: A novel peptide with antimicrobial and anticancer activities from the skin secretion of the south american orange-legged leaf frog, *Pithecopus (Phyllomedusa) hypochondrialis*. *Molecules* 22. doi:10.3390/molecules22101805

Hwang, I.-s., Hwang, J.-s., Hwang, J. H., Choi, H., Lee, E., Kim, Y., et al. (2013). Synergistic effect and antibiofilm activity between the antimicrobial peptide coprisin and conventional antibiotics against opportunistic bacteria. *Current Microbiology* 66, 56–60. doi:10.1007/s00284-012-0239-8

Kharidia, R. and Liang, J. F. (2011). The activity of a small lytic peptide PTP-7 on Staphylococcus aureus biofilms. *J Microbiol* 49, 663–668

Li, M., Xi, X., Ma, C., Chen, X., Zhou, M., Burrows, J. F., et al. (2019). A novel dermaseptin isolated from the skin secretion of *Phyllomedusa tarsius* and its cationicity-enhanced analogue exhibiting effective antimicrobial and anti-proliferative activities. *Biomolecules* 9. doi:10.3390/biom9100628

Liu, Y., Wang, L., Zhou, X., Hu, S., Zhang, S., and Wu, H. (2011). Effect of the antimicrobial decapeptide KSL on the growth of oral pathogens and *Streptococcus mutans* biofilm. *Int J Antimicrob Agents* 37, 33–38

Luca, V., Stringaro, A., Colone, M., Pini, A., and Mangoni, M. L. (2013). Esculentin(1-21), an amphibian skin membrane-active peptide with potent activity on both planktonic and biofilm cells of the bacterial pathogen *Pseudomonas aeruginosa*. *Cellular and Molecular Life Sciences* 70, 2773–2786. doi:10.1007/s00018-013-1291-7

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The embl-ebi search and sequence analysis tools api in 2019. *Nucleic acids research* 47, W636—W641. doi:10.1093/nar/gkz268

Mataraci, E. and Dosler, S. (2012). In vitro activities of antibiotics and antimicrobial cationic peptides alone and in combination against methicillin-resistant *Staphylococcus aureus* biofilms. *Antimicrob Agents Chemother* 56, 6366–6371

Molhoek, E. M., van Dijk, A., Veldhuizen, E. J., Haagsman, H. P., and Bikker, F. J. (2011). A cathelicidin-2-derived peptide effectively impairs Staphylococcus epidermidis biofilms. *Int J Antimicrob Agents* 37, 476–479

Nagant, C., Pitts, B., Nazmi, K., Vandenbranden, M., Bolscher, J. G., Stewart, P. S., et al. (2012). Identification of peptides derived from the human antimicrobial peptide Il-37 active against biofilms formed by *Pseudomonas aeruginosa* using a library of truncated fragments. *Antimicrobial Agents and Chemotherapy* 56, 5698–5708. doi:10.1128/AAC.00918-12

Pompilio, A., Scocchi, M., Pomponio, S., Guida, F., Di Primio, A., Fiscarelli, E., et al. (2011). Antibacterial and anti-biofilm effects of cathelicidin peptides against pathogens isolated from cystic fibrosis patients. *Peptides* 32, 1807–1814
Ramamourthy, G., Arias, M., Nguyen, L. T., Ishida, H., and Vogel, H. J. (2019). Expression and purification of chemokine mip-3α (ccl20) through a calmodulin-fusion protein system. *Microorganisms* 7. doi:10.3390/microorganisms7010008

Sánchez-Gómez, S., Ferrer-Espada, R., Stewart, P. S., Pitts, B., Lohner, K., and Martínez de Tejada, G. (2015). Antimicrobial activity of synthetic cationic peptides and lipopeptides derived from human lactoferrin against pseudomonas aeruginosa planktonic cultures and biofilms. *BMC Microbiology* 15, 137. doi:10.1186/s12866-015-0473-x

Schiffer, M. and Edmundson, A. B. (1967). Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophysical Journal* 7, 121–135. doi:10.1016/S0006-3495(67)86579-2

Shang, D., Liang, H., Wei, S., Yan, X., Yang, Q., and Sun, Y. (2014). Effects of antimicrobial peptide L-K6, a temporin-1CEb analog on oral pathogen growth, Streptococcus mutans biofilm formation, and anti-inflammatory activity. *Appl Microbiol Biotechnol* 98, 8685–8695

Singh, H., Singh, S., and Singh Raghava, G. P. (2019). Peptide secondary structure prediction using evolutionary information. *bioRxiv* doi:10.1101/558791

Sullivan, R., Santarpia, P., Lavender, S., Gittins, E., Liu, Z., Anderson, M. H., et al. (2011). Clinical efficacy of a specifically targeted antimicrobial peptide mouth rinse: Targeted elimination of *Streptococcus* mutans and prevention of demineralization. *Caries Research* 45, 415–428. doi:10.1159/000330510

Sutton, J. M. and Pritts, T. A. (2014). Human beta-defensin 3: a novel inhibitor of Staphylococcus-produced biofilm production. Commentary on “Human β-defensin 3 inhibits antibiotic-resistant Staphylococcus biofilm formation”. *J Surg Res* 186, 99–100

Tao, R., Tong, Z., Lin, Y., Xue, Y., Wang, W., Kuang, R., et al. (2011). Antimicrobial and antibiofilm activity of pleurocidin against cariogenic microorganisms. *Peptides* 32, 1748–1754

Wang, H., He, H., Chen, X., Zhou, M., Wei, M., Xi, X., et al. (2020). A novel antimicrobial peptide (kassinatuerin-3) isolated from the skin secretion of the african frog, *kassina senegalensis*. *Biology (Basel)* 9

Wang, W., Tao, R., Tong, Z., Ding, Y., Kuang, R., Zhai, S., et al. (2012). Effect of a novel antimicrobial peptide chrysophsin-1 on oral pathogens and Streptococcus mutans biofilms. *Peptides* 33, 212–219

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi:10.1093/bioinformatics/btp033

Wei, G.-X., Campagna, A. N., and Bobek, L. A. (2006). Effect of muc7 peptides on the growth of bacteria and on streptococcus mutans biofilm. *Journal of Antimicrobial Chemotherapy* 57, 1100–1109. doi:10.1093/jac/dkl120

Xu, W., Zhu, X., Tan, T., Li, W., and Shan, A. (2014). Design of embedded-hybrid antimicrobial peptides with enhanced cell selectivity and anti-biofilm activity. *PLOS ONE* 9, 1–13. doi:10.1371/journal.pone.0098935

Zhang, R., Zhou, M., Wang, L., McGrath, S., Chen, T., Chen, X., et al. (2010). Phylloseptin-1 (psn-1) from phyllomedusa sauvagei skin secretion: a novel broad-spectrum antimicrobial peptide with antibiofilm activity. *Molecular Immunology* 47 11-12, 2030–7