



AIPpred: Sequence-Based Prediction of Anti-inflammatory Peptides Using Random Forest

Balachandran Manavalan^{1*}, Tae H. Shin^{1,2}, Myeong O. Kim³ and Gwang Lee^{1,2*}

¹ Department of Physiology, Ajou University School of Medicine, Suwon, South Korea, ² Institute of Molecular Science and Technology, Ajou University, Suwon, South Korea, ³ Division of Life Science and Applied Life Science (BK21 Plus), College of Natural Sciences, Gyeongsang National University, Jinju, South Korea

OPEN ACCESS

Edited by:

Jianfeng Pei,
Peking University, China

Reviewed by:

William Farias Porto,
Universidade Católica Dom Bosco,
Brazil
Hao Lin,
University of Electronic Science
and Technology of China, China

*Correspondence:

Balachandran Manavalan
bala@ajou.ac.kr
Gwang Lee
glee@ajou.ac.kr

Specialty section:

This article was submitted to
Experimental Pharmacology and Drug
Discovery,
a section of the journal
Frontiers in Pharmacology

Received: 09 January 2018

Accepted: 12 March 2018

Published: 27 March 2018

Citation:

Manavalan B, Shin TH, Kim MO and
Lee G (2018) AIPpred:
Sequence-Based Prediction
of Anti-inflammatory Peptides Using
Random Forest.
Front. Pharmacol. 9:276.
doi: 10.3389/fphar.2018.00276

The use of therapeutic peptides in various inflammatory diseases and autoimmune disorders has received considerable attention; however, the identification of anti-inflammatory peptides (AIPs) through wet-lab experimentation is expensive and often time consuming. Therefore, the development of novel computational methods is needed to identify potential AIP candidates prior to *in vitro* experimentation. In this study, we proposed a random forest (RF)-based method for predicting AIPs, called AIPpred (AIP predictor in primary amino acid sequences), which was trained with 354 optimal features. First, we systematically studied the contribution of individual composition [amino acid-, dipeptide composition (DPC), amino acid index, chain-transition-distribution, and physicochemical properties] in AIP prediction. Since the performance of the DPC-based model is significantly better than that of other composition-based models, we applied a feature selection protocol on this model and identified the optimal features. AIPpred achieved an area under the curve (AUC) value of 0.801 in a 5-fold cross-validation test, which was ~2% higher than that of the control RF predictor trained with all DPC composition features, indicating the efficiency of the feature selection protocol. Furthermore, we evaluated the performance of AIPpred on an independent dataset, with results showing that our method outperformed an existing method, as well as 3 different machine learning methods developed in this study, with an AUC value of 0.814. These results indicated that AIPpred will be a useful tool for predicting AIPs and might efficiently assist the development of AIP therapeutics and biomedical research. AIPpred is freely accessible at www.thegleelab.org/AIPpred.

Keywords: AIPpred, anti-inflammatory peptides, random forest, hybrid features, parameter optimization

INTRODUCTION

Inflammatory responses are tightly controlled under normal conditions and are essential for the initiation of protective immunity (Medzhitov, 2008; Basith et al., 2011b, 2012). When these responses occur in the absence of infection or persist after their routine function, these processes become pathological, resulting in chronic inflammation and autoimmune disorders, including neurodegenerative disease, rheumatoid arthritis, asthma, psoriasis, diabetes, and multiple sclerosis (Asadullah et al., 2002; Balague et al., 2009; Murdoch and Lloyd, 2010; Steinman et al., 2012; Patterson et al., 2014). The current therapy for inflammatory and autoimmune disorders involves

the use of non-specific anti-inflammatory drugs and other immunosuppressants, which are often associated with potential side effects, such as ineffectiveness against inflammatory disorders and induction of a higher risk of infectious diseases (Tabas and Glass, 2013).

Because peptide-based therapy has several advantages over small molecules owing to their high specificity and minimal toxicity under normal conditions, anti-inflammatory peptides (AIPs) act as potent therapeutic agents for inflammatory and autoimmune disorders (de la Fuente-Nunez et al., 2017; Wu et al., 2017). For example, chronic nasal administration of human amyloid- β peptide (40 amino acid residues) in an Alzheimer's disease mouse model resulted in reduced deposition of amyloid- β , which is a pathological marker of Alzheimer's disease, microgliosis, astrocytosis, and neuritic dystrophy in the brain (Weiner et al., 2000). Vasoactive intestinal peptide reduces inflammation in rheumatoid arthritis by altering the immune response to reduce cytokine production in CD4⁺ T cells (Delgado et al., 2001). RDP58, a synthetic decapeptide, effectively inhibits the production of inflammatory cytokines, such as tumor necrosis factor- α , interferon (IFN)- γ , IL-2, and IL-12, as well as the infiltration of inflammatory cells associated with urothelial inflammatory response in an *in vivo* model of lipopolysaccharide-induced cystitis (Boismenu et al., 2002; Gonzalez et al., 2005). Furthermore, AIPs act as potent candidates for cancer prevention and therapy because inflammation is closely linked to cancer (Rayburn et al., 2009).

Although AIPs specifically bind to the receptor and activate signaling cascades in cells, experimental identification and development of novel AIPs represent extremely expensive and often time-consuming processes. Therefore, the development of sequence-based computational methods is necessary to allow the rapid identification of potential AIP candidates prior to their synthesis. It should be noted that the prediction methods prior to synthesis would help a number of previous design studies (Geetha et al., 2005; Grieco et al., 2005; Park et al., 2009). To this end, Gupta et al. (2017b) developed a support vector machine (SVM)-based method to predict AIPs using trinucleotide composition and motif features. This represents the first and only method available for AIP prediction, and although this method has stimulated further development in this area, additional work is needed for the following reasons: (i) with the steadily increasing number of anti-inflammatory epitopes or peptides in the Immune Epitope Database (IEDB), it is necessary to develop more accurate prediction methods with a larger benchmark dataset. (ii) The feature space used by the existing method is incomplete; hence, additional potent features are needed to be characterized. Owing to these deficiencies, other methods are necessitated to accurately predict AIPs by taking advantage of machine learning (ML) algorithms and informative feature extraction based on high-quality benchmarking datasets.

In this study, we developed a random forest (RF)-based method to predict AIPs, called AIPred (AIP predictor from primary amino acid sequences), in which optimal features were selected using a feature selection protocol, which has been implemented in addressing various biological problems (Manavalan and Lee, 2017; Manavalan et al., 2017b, 2018). First,

we studied the contribution of individual composition [amino acid composition (AAC), amino acid index (AAI), dipeptide composition (DPC), chain-transition-composition (CTD), and physicochemical properties (PCP)] in AIP prediction. Since the DPC-based model significantly outperformed other composition-based models, we applied a feature selection protocol on DPC and identified the optimal features. In addition to AIPred, we also developed SVM, extremely randomized tree (ERT), and *k*-nearest neighbors (*k*-NN)-based methods. It is to be noted that, when objectively evaluated using an independent dataset, AIPred displayed superior performance compared to the currently available method AntiInflam and 3 other ML methods (ERT, SVM, and *k*-NN) developed in this study.

MATERIALS AND METHODS

For the development of our method, we followed the 5 guidelines (Chou, 2011) mentioned in a series of recent publications (Chen W. et al., 2016; Chen et al., 2017; Feng et al., 2017; Liu et al., 2017) on new peptide-prediction methods that could be easily accessed by both experimentalists and theoretical scientists: (i) construct a valid benchmarking dataset to train and test the prediction model; (ii) formulate the biological-sequence samples with an effective mathematical expression truly reflecting their intrinsic correlation with the target to be predicted; (iii) introduce or develop a powerful algorithm (or engine) to operate the prediction; (iv) properly perform cross-validation tests to objectively evaluate the anticipated accuracy of the predictor; and (v) establish a user-friendly web server for the predictor that is accessible to the public. Below, we describe in detail the application of each of these steps.

Dataset Construction

To build a classification model, a well curated dataset is required. Hence, we extracted experimentally validated positive and negative linear peptides or epitopes from the IEDB (Zhang et al., 2008; Fleri et al., 2017). A peptide induced any one of the anti-inflammatory cytokines [IL-10, IL-4, IL-13, IL-22, TGF β , and IFN- α/β] in T-cell assays of human and mouse (Marie et al., 1996), was considered positive. Similarly, linear peptides testing negative for anti-inflammatory cytokines were considered negative. To generate a non-redundant (nr) dataset, we eliminated redundant peptides using CD-HIT (Huang et al., 2010) by applying a sequence identity threshold of 0.8, indicating that sequence identity between any two sequences greater than 80% is discarded. Using a more stringent criterion, such as 30 or 40%, as imposed in (Gupta et al., 2013; Ding et al., 2014; Chen X-X. et al., 2016), could improve the credible reliable of the model. However, in this study we do not use such a stringent criterion, because the currently available data does not allow it. Otherwise, the number of samples for some subsets would be insufficient for statistical significance.

Finally, we obtained an nr dataset of 1678 AIPs and 2,516 non-AIPs, whose size is \sim 2-fold bigger than the dataset used in the previous method (i.e., AntiInflam) (Gupta et al., 2017b). From this nr dataset, 80% of the data was randomly selected

as the benchmarking dataset (i.e., 1258 AIPs and 1,887 non-AIPs) to develop a prediction model, whereas the remaining 20% was considered the independent dataset (i.e., 420 AIPs and 629 non-AIPs).

Feature Extraction

We formulated the AIP-prediction task as a binary classification problem (AIP or non-AIP) and solved it using RF, SVM, *k*-NN, and ERT algorithms. An important aspect of this process involves the extraction of a set of relevant features. Therefore, we used AAC, AAI, DPC, PCP, and CTD, whose definitions are briefly discussed in the following subsections.

Amino Acid Composition

AAC is defined as the fraction of each amino acid in the given peptide sequence, and it was calculated using the following equation (1).

$$\text{AAC}(i) = \frac{\text{Frequency of amino acid } (i)}{\text{Length of the peptide}} \quad (1)$$

where *i* can be any one of the 20 natural amino acids. AAC has a fixed length of 20 features.

Amino Acid Index

The AAIndex database contains amino acid indices of various physicochemical and biochemical properties (Kawashima et al., 2008). Saha et al. (2012) classified these amino acid indices into eight clusters, and the central indices of each cluster were named as high-quality amino acid indices: BLAM930101, BIOV880101, MAXF760101, TSAJ990101, NAKH920108, CEDJ970104, LIFS790101, and MIYS990104. We averaged eight high-quality amino acid indices (i.e., a 20-dimensional vector) as an input feature.

CTD

The CTD feature was introduced by Dubchak et al. (1995) for predicting protein-folding classes. Thereafter, it was successfully applied in various sequence-based classification algorithms (Cai et al., 2003; Magnan et al., 2009; Wang et al., 2016; Hasan et al., 2017). CTD represents the distribution of amino acid patterns along the primary sequence, based on their physicochemical or structural properties. There are seven physicochemical properties, including hydrophobicity, polarizability, normalized van der Waals volume, secondary structure, polarity, charge and solvent accessibility.

All amino acids are divided into three groups: polar, neutral and hydrophobic. C consists of three percentage composition values for a given peptide: polar, neutral and hydrophobic. T consists of the percentage frequency of a polar followed by a neutral residue or of a neutral by a polar residue. It may also consist of a polar, followed by a hydrophobic residue or a hydrophobic followed by a polar residue. It may also consist of a neutral, followed by a hydrophobic or a hydrophobic, followed by a neutral residue. D consists of five values for each of the three groups. It measures the chain length, within which the first, 25, 50, 75, and 100 % of the amino acids of a specific property are located. There are three descriptors and

$3(C) + 3(T) + 5 \times 3(D) = 21$ descriptor values for a single amino acid attribute. Consequently, seven different amino acid attributes produce a total of $7 \times 21 = 147$ features.

Dipeptide Composition

DPC is defined as the total number of dipeptides normalized against 400 possible dipeptides in the given peptide sequence and was calculated using the following equation (2):

$$\text{DPC}(i) = \frac{\text{Total number of dipeptides } (i)}{\text{Total number of all possible dipeptides}'} \quad (2)$$

where *i* can be any one of the 400 possible dipeptides. DPC has a fixed length of 400 features.

Physicochemical Properties

Frequencies of the following features are directly computed from the sequence consisting of: (1) hydrophobic (i.e., F, I, W, L, V, M, Y, C, A); (2) hydrophilic (i.e., R, K, N, D, E, P); (3) neutral (i.e., T, H, G, S, Q); (4) positively charged (i.e., K, H, R); (5) negative-charged (i.e., D, E); (6) turn-forming residues fraction (i.e., $(N + G + P + S)/n$, where *n* = sequence length); (7) absolute charge per residue (i.e., $|\frac{R+K-D-E}{n} - 0.03|$); (8) molecular weight and (9) aliphatic index (i.e., $(A+2.9V+3.9I+3.9L)/n$).

Machine Learning Methods

In general, the major advantage of the ML method is that it can identify the hidden relationship between the input features and the objective values in a complex dataset, which will be helpful for accurate prediction (Cao et al., 2014, 2016a,b, 2017; Manavalan et al., 2014, 2017a, 2010b; Cao and Cheng, 2016; Manavalan and Lee, 2017). In this study, we used 4 different ML methods (ERT, RF, *k*-NN and SVM) to develop their prediction models using benchmarking datasets. The description of these methods is provided as follows.

Random Forest

Breiman (2001) proposed RF as an ensemble technique to perform predictions using 100s or 1000s of independent decision trees. RF is one of the most popular ML methods and is used as a computational approach to numerous biological problems. Detailed descriptions of the RF algorithm have been provided in earlier studies (Lee et al., 2013, 2015; Manavalan et al., 2014). In the RF algorithm, the number of trees (*n*tree), variables randomly chosen at each node split (*m*try), and the minimum number of samples required to split an internal node (*nsplit*) are the 3 most influential parameters that require optimization. We optimized these parameters using a grid search within the following ranges: *n*tree from 50 to 1,000, with a step size of 20; *m*try from 1 to 7, with a step size of 1; and *nsplit* from 2 to 10, with a step size of 1.

Extremely Randomized Tree

Geurts et al. (2006) proposed ERT as an ensemble technique utilizing hundreds of independent decision trees to perform classification. Although the ERT algorithm is similar to that of RF, the major differences are that ERT uses the entire training sample instead of a bootstrap sample (RF) to construct a tree, and the ERT splitting criterion is random, whereas RF uses information

gain measured by the Gini impurity. Furthermore, the parameter-optimization procedure is the same as that used for the RF method.

Support Vector Machine and k -Nearest Neighbors

Descriptions of SVM and k -NN along with their optimization procedures have been provided in earlier studies (Manavalan et al., 2015, 2017a, 2010b; Manavalan and Lee, 2017). We followed the same procedures in this study.

Evaluation Metrics

To compare the prediction methods, we used the following five metrics: sensitivity, specificity, accuracy, Mathews' correlation coefficient (MCC), and the area under receiver operating characteristics (ROC). All these metrics are commonly used in the literature to measure the quality of binary classification (Porto et al., 2017b).

$$\left\{ \begin{array}{l} \text{Sensitivity} = \frac{TP}{PS}, \\ \text{Specificity} = \frac{TN}{NS}, \\ \text{Accuracy} = \frac{TP+TN}{PS+NS}, \\ \text{MCC} = \frac{1 - \left(\frac{FN}{PS} + \frac{FP}{NS}\right)}{\sqrt{\left(1 + \frac{FP-FN}{PS}\right)\left(1 + \frac{FN-FP}{NS}\right)}}, \end{array} \right. \quad (3)$$

where TP, FN, TN, and FP respectively represent the number of true positive, false negative, true negative and false positive. PS and NS respectively represent the total number of sequences in the positive set (AIPs) and negative set (non-AIPs).

AUC is the area under the ROC curve, representing the relationship between TP rate and FP rate of the model. The AUC is an indicator of the performance quality of the binary classifier.

Development of a Prediction Server

We developed an online prediction server using hypertext mark-up language and JavaScript, with a Python script executed in the background upon submission of peptide sequences in the FASTA format. Users can submit single or multiple sequences containing only standard amino acid residues in FASTA format. The AIPpred web server outputs the predicted class along with probability values for the given peptide sequence.

RESULTS

Compositional and Positional Information Analysis

We performed compositional analysis using the combined dataset (i.e., benchmarking and independent). AAC analysis revealed that average composition of certain residues, including Arg, Leu and Lys, were dominant in AIPs. However, Gly, Asp, Val, Tyr and Pro were dominant in non-AIPs (Welch's t -test; $P \leq 0.05$) (Figure 1A). Furthermore, DPC analysis revealed that 19% of dipeptides differed significantly between AIPs and non-AIPs (Welch's t -test; $P \leq 0.05$). Of these, the top-10 most abundant dipeptides in AIPs and non-AIPs were LL, SL, LE, LI, LS, LK, YL, IK, RI and KR, and DV, KG, DD, EF, GD, FD, YP, TY, GH and HV, respectively (Figure 1B). These results

suggest that the most abundant dipeptides in AIPs consist primarily of pairs of aliphatic-aliphatic, positively charged-positively charged or -aliphatic, and hydroxyl group-aliphatic or -aromatic amino acids. However, the most abundant dipeptides in the non-AIPs were negatively charged-negatively charged or -aliphatic, and positively charged-negatively charged amino acids. Overall, significant differences observed in compositional analysis could be incorporated into ML algorithms to improve prediction performances. Hence, we considered them as input features.

To understand the positional information of each residue, a sequence logo of the first ten residues from the N- and the C-terminal of AIPs and non-AIPs were generated using two sample logos. To test their statistical significance, the height of the peptide logos were scaled (t -test by $P < 0.05$). At the N-terminal, we found that, compared to other amino acids, R, at positions 4 and 5; L, at positions 1, 5, 7, 8, and 10; F, at positions 4 and 8; and Y, at positions 4 and 6 were significantly overrepresented. Alternatively, negatively charged residue D, at positions 4, 8, and 10; and S/T, at positions 3, 6, and 7 were significantly underrepresented (Figure 1C). No significant amino acids were found at enriched position 9 or the depleted positions 2 and 9. C-terminal R/K, at positions 2 and 10; and L, at positions 2, 5, 6, 7, and 8 were significantly overrepresented. Alternatively, negatively charged residues D, at positions 3, 5, and 8 and G, at positions 4 and 5 were significantly underrepresented (Figure 1D). No significant amino acids were found at enriched position 1, 3, 4, and 9 or the depleted positions 1, 6, and 9. These results suggest that comparatively residues, L and R/K, are preferred in AIPs. This is consistent with the AAC analysis observation. Furthermore, positional preference analysis will be helpful for experimenters who design *de novo* AIPs and substitute amino acids at particular positions to make the peptides more effective.

The Overall Framework of the AIPpred Approach

The overall framework of AIPpred is shown in Figure 2. It consists of the following 4 stages: (1) construction of a nr benchmarking dataset of 3,145 peptides (1,258 AIPs and 1,887 non-AIPs) and an independent dataset of 1,049 peptides (420 AIPs and 629 AIPs); (2) extraction of various features from peptide sequences, including AAC, AAI, CTD, DPC, and PCP; (3) systematic evaluation of individual composition and generation of 35 different feature sets based on the feature importance scores (FISs) computed using the RF algorithm. These different feature sets were inputted to RF, and their respective prediction models were built; and (4) selection of the best model.

Performances of RF Models Based on Individual Composition

To test the effectiveness of individual composition in AIP prediction, we inputted each composition separately to RF and developed their corresponding prediction models, as well as a model based on hybrid features (linear combination of individual composition). The performance of these models is shown in

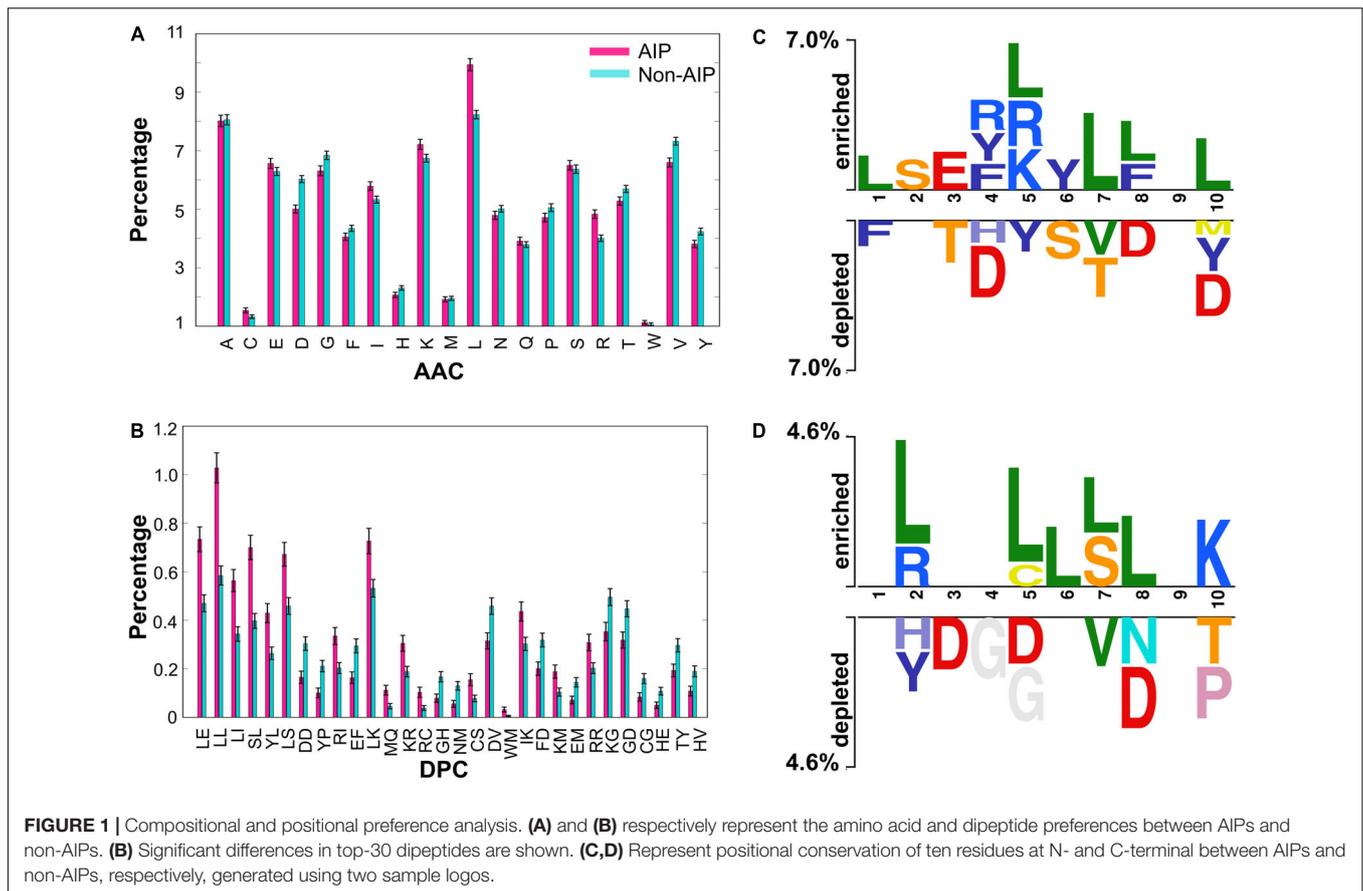


Figure 3. At a P -value threshold of 0.05, the DPC-based model significantly outperformed 4 other individual (PCP, AAI, CTD, and AAC) composition-based models and hybrid (H) models. Hence, we considered only the DPC-based model for further analysis. In the DPC-based model, all possible dipeptides are not equally important for the trained model performance. The inclusion of irrelevant dipeptides during training might reduce model performance. Therefore, a feature selection paradigm is essential to remove irrelevant dipeptides and consequently improve the prediction performance.

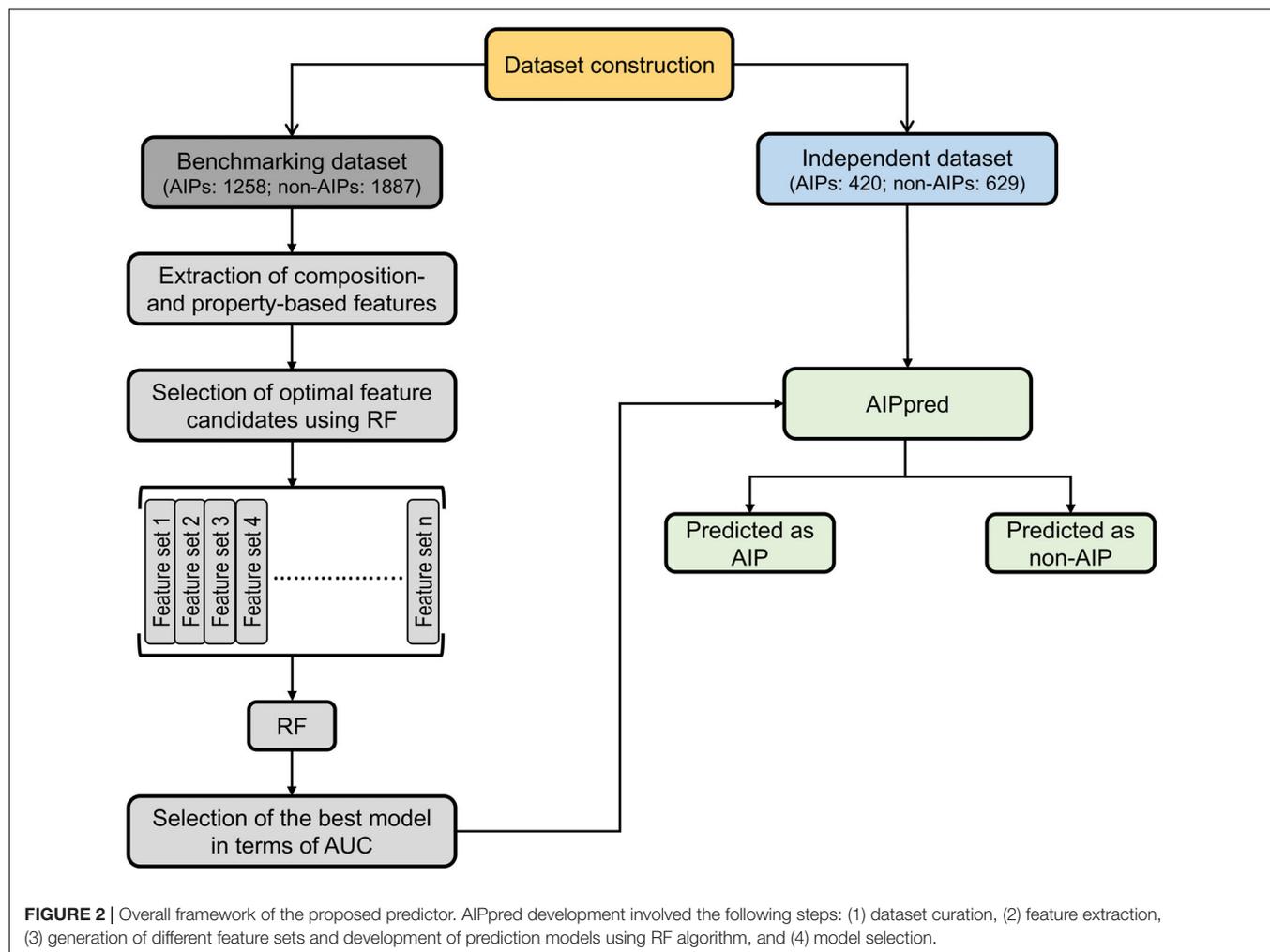
Feature Selection Protocol

The feature selection protocol employed in this study is the same as the one used in recent studies (Manavalan et al., 2017b; Manavalan and Lee, 2017). First, we applied the RF algorithm and estimated the FISs of 400 dipeptides in distinguishing AIPs and non-AIPs. In short, all features were inputted to the RF, and 5-fold cross-validation was carried out using the benchmarking dataset. For each round of cross-validation, we built 10,000 trees, and the number of variables at each node was chosen randomly from 1 to 50. The average FISs from all the trees are shown in **Figure 4A** and Supplementary Table S1, where $\sim 36\%$ of the features ($FIS \geq 0.003$) contributed significantly to AIP prediction. Second, we excluded 9 features that have a low FIS (less than 0.0005) and generated 35 different feature sets based on FIS cut-off ($0.0005 \leq FIS \leq 0.0039$, with a step size of

0.0001) with the remaining 391 features. In general, the optimal feature set lies in between a large number of features that contain considerable irrelevant information and a small number of only important features (describing a part of AIP properties). The 35 different feature sets generated have a feature size ranging from 49 to 391. Basically, we eliminated less important features in a step-wise manner. Finally, we inputted each set into the RF algorithm and optimized ML parameters (*mtry*, *ntree*, and *nsplit*) by 5-fold cross-validation on the benchmarking dataset. To check the robustness of the model performance, we carried out 5-fold cross-validation 10 times by randomly portioning the benchmarking dataset and considering median ML parameters and average performance measures. Finally, the performances of 35 prediction models were compared, and the best model that produced the highest area under the curve (AUC), whose corresponding feature set was considered optimal, was selected.

Selection of the Optimal Model

Figure 4B shows the performances of the RF-based models in terms of AUC using different feature sets, where a fluctuation was found in the initial phase, peaking in an F354-based model with an AUC of 0.801. Afterward, the AUC showed a stable performance followed by downward trend with the decrease in the number of features. Here, we selected the F354-based model as the final one owing to its best performance and named it AIPred; its optimal ML parameters were *ntree* = 430, *mtry* = 1,



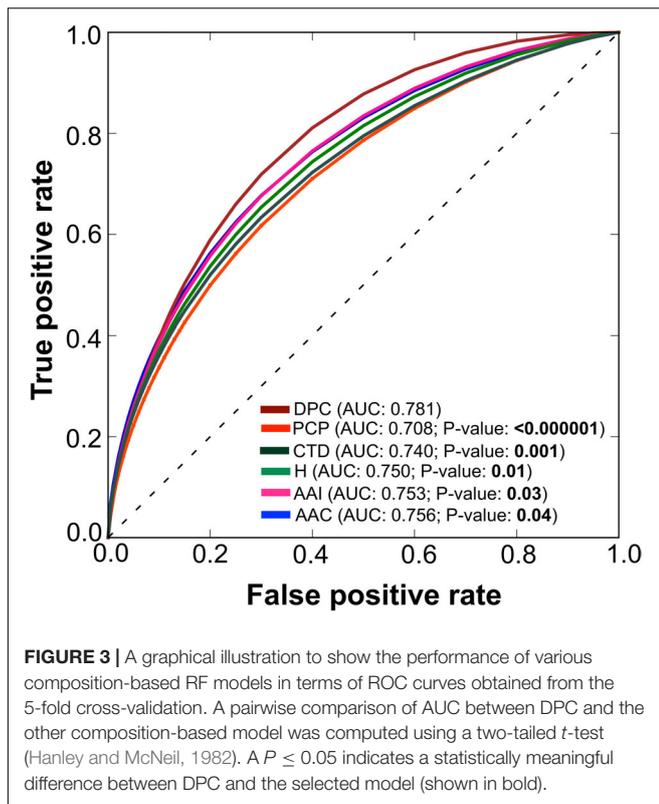
and $nsplit = 2$. Interestingly, our feature selection protocol excluded most of the Trp, Cys and Met containing dipeptides and selected the remaining 354 dipeptides as optimal candidates that covered all 20 amino acids (Supplementary Table S1), which produced the best performance.

Due to the imbalanced dataset, the optimal probability cut-off value of 0.36 was chosen via grid search for AIPred to define the class. To demonstrate the effect of our feature selection protocol, we compared AIPred with the control (using all DPC features). **Figure 4C** shows that AIPred MCC, accuracy, and AUC were respectively 5, 2.5, and ~4% higher than those of the control. These results demonstrated that the numerous redundant or uninformative features present in the original feature set were eliminated through our feature selection protocol, thereby significantly improving the performance.

Comparison of AIPred With Other ML Algorithms

Generally, ML-based methods are problem specific (Dreiseitl et al., 2001; Silva et al., 2011; Khondoker et al., 2016). Hence, it is necessary to explore different ML methods on the same

dataset to select the best one instead of selecting a ML method arbitrarily. In addition to RF, we also developed ERT-, k -NN-, and SVM-based models using the same feature selection protocol and benchmarking dataset. Each ML method has its own advantages and disadvantages (Khan et al., 2010). A detailed description of these 3 methods has been provided in our recent studies (Manavalan and Lee, 2017; Manavalan et al., 2017b). Here, the procedure of ML parameter optimization for these 3 methods, final model selection, and optimal probability cut-off value was the same as that for AIPred. The overall performance comparison of the RF method with the other 3 methods is shown in **Figure 5**, where RF and ERT produced a similar performance regardless of the feature set used, thus indicating that ensemble-based algorithm is better suited for AIP prediction. Interestingly, the final selected model for the 3 methods (SVM, ERT, and k -NN) is better than that of their corresponding control (using all dipeptide composition), again emphasizing the efficiency of the feature selection protocol. We also checked whether the final selected optimal model for these 3 methods is better than other composition-based and hybrid models. **Figure 6** shows that the optimal model significantly better than their counterparts.



Finally, we compared AIPred performance with that of the other 3 methods; the results are shown in **Table 1**, where the methods are ranked according to the AUC associated with predictive capability. The accuracy, AUC, and MCC of AIPred were higher than those of other methods by 0.5–9%, 0.6–11%, and 1–17%, respectively. Using a *P*-value threshold of 0.05, AIPred significantly outperformed SVM and *k*-NN, and was better than ERT, thus indicating the superiority of AIPred. To check the transferability of AIPred, we evaluated

an independent dataset and compared it with the state-of-the-art method and 3 other ML methods developed in this study.

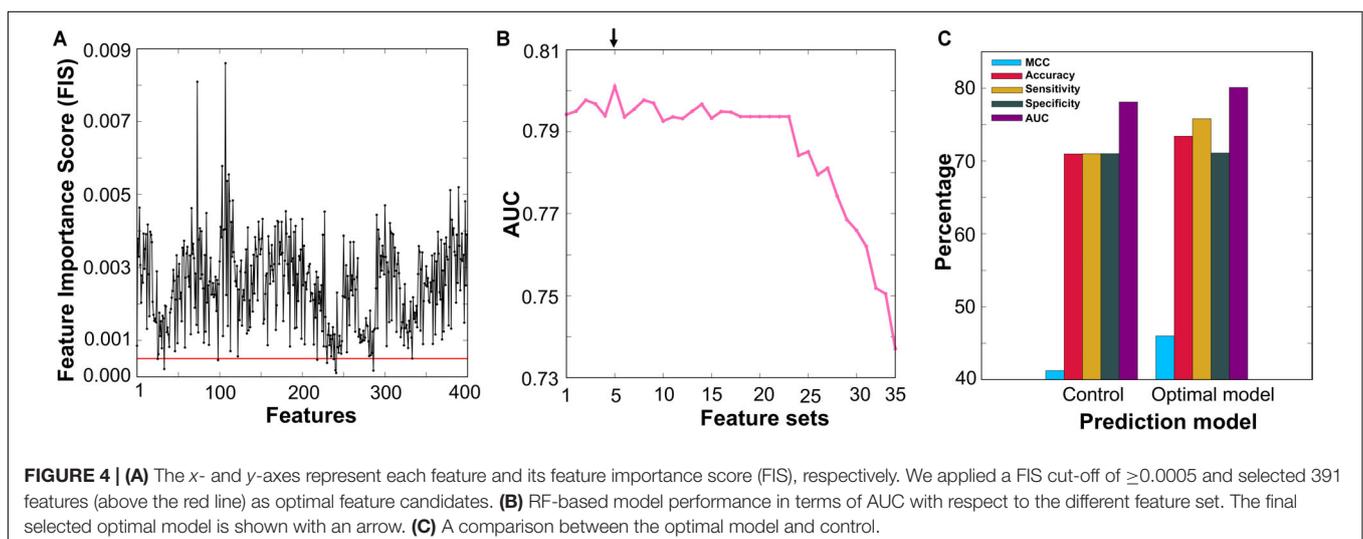
Performance of Various Methods on an Independent Dataset

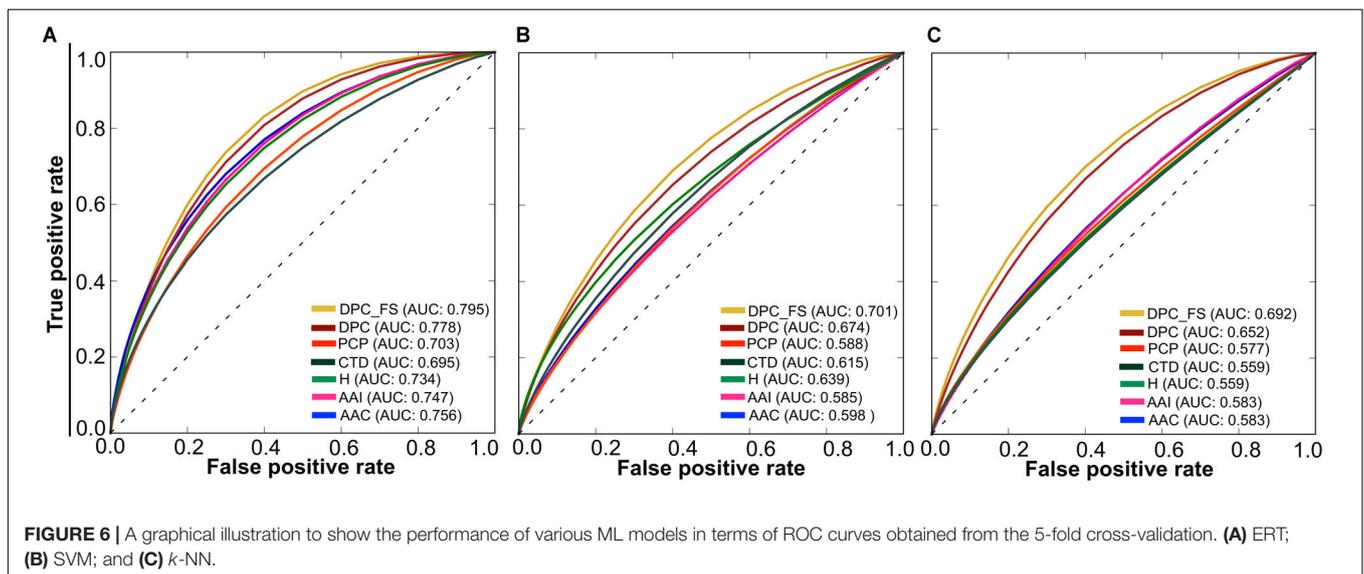
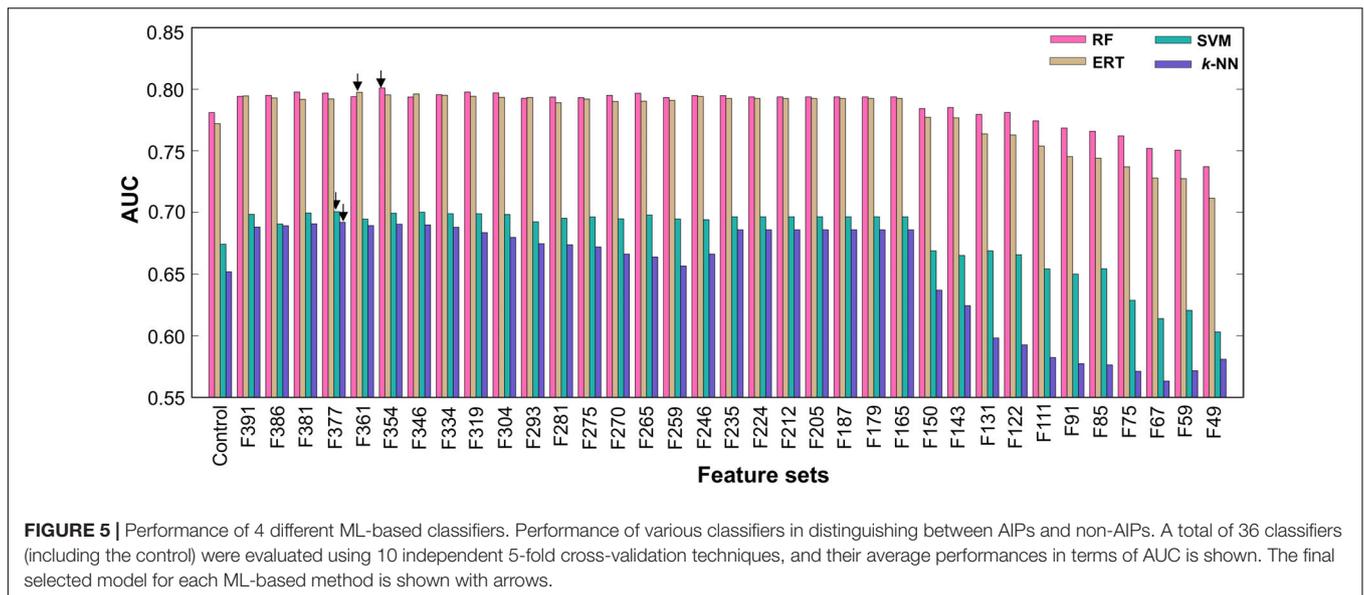
We evaluated the performances of our 4 methods along with that of the state-of-the-art method (AntiInflam) on an independent dataset. **Table 2** shows that AIPred achieving values of 0.479, 0.744 for MCC and accuracy, respectively. Indeed, the corresponding metrics were ~2–28% and ~1–17%, higher than those achieved by other methods, indicating superiority of AIPred.

Using a *P*-value threshold of 0.05, AIPred significantly outperformed SVM, *k*-NN and AntiInflam suggesting its usefulness as an improvement to existing tools for predicting AIPs. Interestingly, AIPred performed consistently well, both in training and on an independent dataset (**Figure 7**), suggesting its ability to do well in unseen peptides when compared to other ML-based models developed in this study.

Comparison of AIPred With the AntiInflam Method in Terms of Methodology

A detailed comparison of the differences between AIPred and AntiInflam (Gupta et al., 2017b) in terms of methodology resulted in the following findings: (i) larger size of the benchmarking dataset used to develop AIPred than AntiInflam. (ii) AntiInflam uses an SVM-based algorithm, whereas we explored 4 different ML-based algorithms, including SVM, and reported that the RF-based method produced the best performance, thus making AIPred the first application of an RF-based method in AIP prediction. (iii) AntiInflam uses hybrid features, whereas AIPred uses optimal DPC features identified by the feature selection protocol. (iv) AIPred used a unique parameter-optimization procedure involving 10 independent 5-fold cross-validations to finalize the ML





parameters, whereas only one 10-fold cross-validation was employed for AntiInflam.

The AIPred Online Prediction Server

Prediction methodologies available on a web server are practically beneficial to experimentalists, as well as to developers (Chen et al., 2013, 2017; Chen W. et al., 2016; Liu et al., 2017). A few examples of bioinformatics web servers that have been utilized for protein function predictions are available in the literature (Govindaraj et al., 2010, 2011; Manavalan et al., 2010a,b, 2011; Basith et al., 2013, 2011a). We developed an online prediction server called AIPred.¹ For checking the reproducibility of our findings, the datasets used in this study can be downloaded from the AIPred web server.

¹www.thegleelab.org/AIPred

DISCUSSION

Identifying the peptides that induce anti-inflammatory cytokines is one of the challenging task in the field of vaccine design. The computational identification of AIP candidates is essential for shortening the laborious experimental tasks. AIPs prediction is more challenging than other peptide-based prediction methods, including anticancer, antiviral and cell-penetrating peptides (Thakur et al., 2012; Tang et al., 2016; Manavalan et al., 2017a). All these methods were developed on smaller dataset with negative examples taken from randomly chosen UniProt peptides, which are not experimentally validated. However, we have used experimentally verified AIPs and non-AIPs from IEDB, whose size was ~2-fold bigger than the dataset used in the state-of-the-art method (Gupta et al., 2017b). In general, methods developed using such experimentally verified larger dataset have

TABLE 1 | A Comparison of AIPred performance with other ML-based methods developed in this study using the same benchmarking dataset.

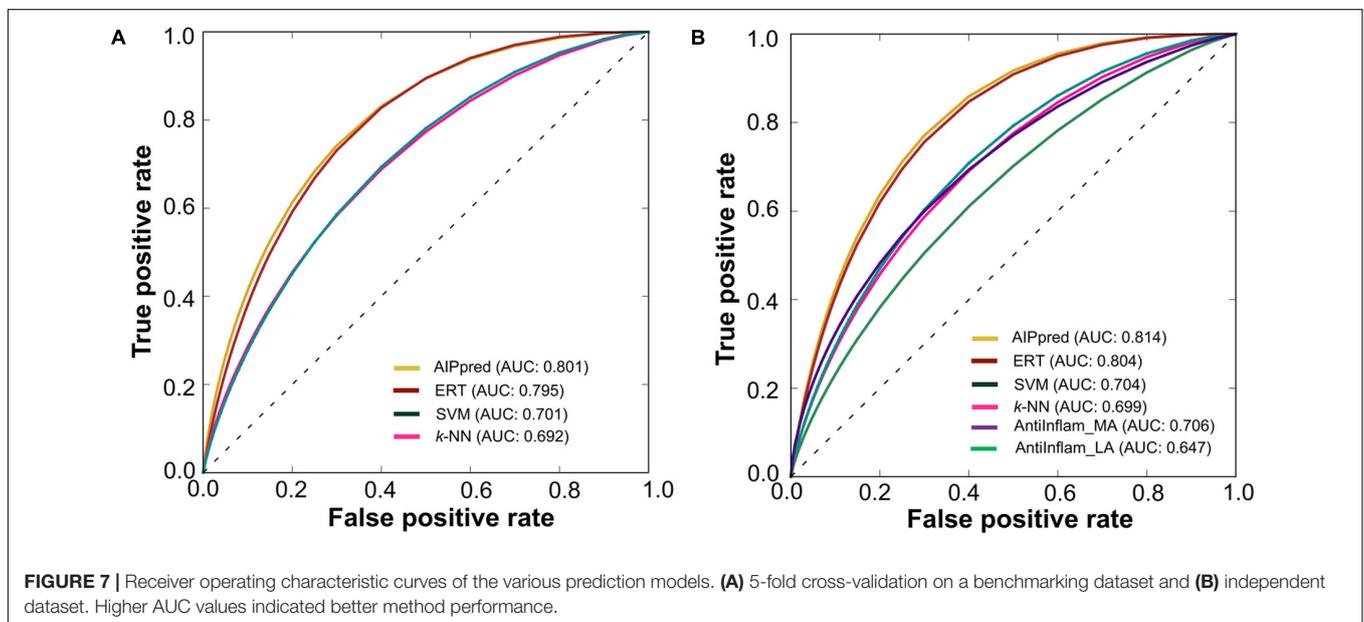
Method	MCC	Accuracy	Sensitivity	Specificity	AUC	P-value
AIPred	0.460	0.734	0.758	0.711	0.801	
ERT	0.451	0.730	0.734	0.726	0.795	0.615
SVM	0.311	0.656	0.642	0.675	0.701	<0.000001
k-NN	0.291	0.641	0.512	0.770	0.692	<0.000001

The first column represents the method developed in this study. The second, the third, the fourth, and the fifth respectively represent the MCC, accuracy, sensitivity, and specificity. The sixth column and the seventh represent the area under curve (AUC) and pairwise comparison of AUC between AIPred and the other ML-based methods computed using a two-tailed t-test (Hanley and McNeil, 1982). A $P \leq 0.05$ indicates a statistically meaningful difference between AIPred and the selected method (shown in bold).

TABLE 2 | Performance of various methods on independent dataset.

Method	MCC	Accuracy	Sensitivity	Specificity	AUC	P-value
AIPred	0.479	0.744	0.741	0.746	0.814	
ERT	0.463	0.736	0.731	0.740	0.804	0.621
Antinflam (MA)	0.210	0.601	0.786	0.417	0.706	<0.000001
SVM	0.298	0.651	0.621	0.680	0.704	<0.000001
k-NN	0.296	0.640	0.479	0.801	0.699	<0.000001
Antinflam (LA)	0.197	0.575	0.258	0.892	0.647	<0.000001

The first column represents the method employed in this study. The second, the third, the fourth, and the fifth respectively represent the MCC, accuracy, sensitivity, and specificity. The sixth column and the seventh represent the AUC and pairwise comparison of AUC between AIPred and the other methods computed using a two-tailed t-test. A $P \leq 0.05$ indicates a statistically meaningful difference between AIPred and the selected method (shown in bold). In the first column, LA and MA respectively correspond to less accurate and more accurate prediction method. We note that Antinflam LA and MA classification accuracy was computed using default threshold value of 0.5 and -0.3 (reported in Gupta et al., 2017b), respectively.



a wide range of applications in modern biology (Porto et al., 2017a).

We have made a systematic attempt to understand the nature of anti-inflammatory inducing peptides and to develop the prediction model. The construction of experimentally validated nr dataset is the backbone of this study. We analyzed these peptides to understand the compositional and positional preferences of residues in AIPs, as shown in result section, Leu, Lys and Arg is highly abundant in AIPs,

compared to non-AIPs. Previous studies showed that Leu-Lys rich peptides play an important role in inducing anti-inflammatory cytokines in periodontal disease (Shang et al., 2014). Furthermore, determining the biological significance of various dipeptides in anti-inflammatory induction, observed in our study, requires further studies and experimental validation.

We explored four different ML algorithms (RF, SVM, ERT, and k-NN) and compositional features, including AAC, AAI,

DPC, CTD, and PCP for discriminating AIPs and non-AIPs. It is worth mentioning that all these ML algorithms and five different compositions were used in various sequence-based classification methods (Lata et al., 2007; Dhanda et al., 2013; Gautam et al., 2015; Gupta et al., 2017a; Manavalan et al., 2017a; Nagpal et al., 2017). Since DPC-based model from the respective algorithm produced the best performance among the different compositions, we applied a feature selection protocol on dipeptide composition and selected more important features that further improved the performance. RF produced the best performance among the various ML algorithms and named it AIPpred. Interestingly, our systematic feature selection protocol excluded most of Trp, Cys and Met containing dipeptides and selected the remaining 354 dipeptides as optimal candidates, thus indicating the arrangement of particular local ordering dipeptides plays an important role in AIPs/non-AIPs classification. Furthermore, we demonstrated that AIPpred outperformed a state-of-the-art method (AntiInflam) and 3 other methods (ERT, *k*-NN, and SVM) developed in this study when it was objectively evaluated on an independent dataset. Interestingly, AIPpred performed consistently better in benchmarking and independent datasets, suggesting its ability to predict AIPs of unseen peptides.

The improved performance of AIPpred is mainly due to the following reasons: (i) larger benchmarking dataset utilized for the model development (ii) systematic evaluation of individual composition and the selection of an appropriate composition (i.e., DPC); and applying the feature selection protocol on DPC to select the optimal feature set, which further improves prediction performance; (iii) ML parameters were obtained by a rigorous 5-fold cross-validation procedure. Here, the 5-fold cross-validation procedure was repeated 10 times, with the random portioning of the benchmarking dataset, whose median ML parameters were considered as the final one; and (iv) the choice of ML method. Interestingly, the current approach is a general one, which is applicable to numerous other peptide-based classification problems. Although AIPpred displayed a superior performance over other methods, a pressing need exists for further improvements, incorporating novel features, and exploring different feature selection techniques, including ANOVA (Zhao et al., 2017), *F*-score (Lin et al., 2017), and binomial distribution (Lai et al., 2017).

REFERENCES

- Asadullah, K., Volk, H. D., and Sterry, W. (2002). Novel immunotherapies for psoriasis. *Trends Immunol.* 23, 47–53. doi: 10.1016/S1471-4906(01)02119-6
- Balague, C., Kunkel, S. L., and Godessart, N. (2009). Understanding autoimmune disease: new targets for drug discovery. *Drug Discov. Today* 14, 926–934. doi: 10.1016/j.drudis.2009.07.002
- Basith, S., Manavalan, B., Gosu, V., and Choi, S. (2013). Evolutionary, structural and functional interplay of the IkappaB family members. *PLoS One* 8:e54178. doi: 10.1371/journal.pone.0054178
- Basith, S., Manavalan, B., Govindaraj, R. G., and Choi, S. (2011a). In silico approach to inhibition of signaling pathways of Toll-like receptors 2 and 4 by ST2L. *PLoS One* 6:e23989. doi: 10.1371/journal.pone.0023989

CONCLUSION

The proposed predictor is quite promising in AIP prediction and available as web server at www.thegleelab.org/AIPpred. Even though AIPred represents the second publicly available method for predicting AIPs, the delivery of higher accuracy is noteworthy. Compared to experimental approaches, bioinformatics tools, such as AIPpred represent a powerful and cost-effective approach for proteome-wide prediction of AIPs. Therefore, AIPpred might be useful for large-scale AIP prediction and facilitating hypothesis-driven experimental design.

AUTHOR CONTRIBUTIONS

BM and GL conceived and designed the experiments and wrote the paper. BM performed the experiments. BM and TS analyzed the data. GL and MK contributed reagents/materials/software tools. All authors reviewed the manuscript and agreed to this information prior to submission.

FUNDING

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology [2015R1D1A1A09060192 and 2009-0093826], and the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning [2016M3C7A1904392].

ACKNOWLEDGMENTS

The authors thank Ms. Da Yeon Lee for the secretarial assistance in the preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2018.00276/full#supplementary-material>

- Basith, S., Manavalan, B., Lee, G., Kim, S. G., and Choi, S. (2011b). Toll-like receptor modulators: a patent review (2006-2010). *Expert Opin. Ther. Pat.* 21, 927–944. doi: 10.1517/13543776.2011.569494
- Basith, S., Manavalan, B., Yoo, T. H., Kim, S. G., and Choi, S. (2012). Roles of toll-like receptors in cancer: a double-edged sword for defense and offense. *Arch. Pharm. Res.* 35, 1297–1316. doi: 10.1007/s12272-012-0802-7
- Boismenu, R., Chen, Y., Chou, K., El-Sheikh, A., and Buelow, R. (2002). Orally administered RDP58 reduces the severity of dextran sodium sulphate induced colitis. *Ann. Rheum. Dis.* 61(Suppl. 2), ii19–ii24. doi: 10.1136/ard.61.suppl_2.ii19
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

- Cai, C. Z., Han, L. Y., Ji, Z. L., Chen, X., and Chen, Y. Z. (2003). SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res.* 31, 3692–3697. doi: 10.1093/nar/kg660
- Cao, R., Adhikari, B., Bhattacharya, D., Sun, M., Hou, J., and Cheng, J. (2017). QAcon: single model quality assessment using protein structural and contact information with machine learning techniques. *Bioinformatics* 33, 586–588. doi: 10.1093/bioinformatics/btw694
- Cao, R., Bhattacharya, D., Hou, J., and Cheng, J. (2016a). DeepQA: improving the estimation of single protein model quality with deep belief networks. *BMC Bioinformatics* 17:495.
- Cao, R., and Cheng, J. (2016). Protein single-model quality assessment by feature-based probability density functions. *Sci. Rep.* 6:23990. doi: 10.1038/srep23990
- Cao, R., Jo, T., and Cheng, J. (2016b). Evaluation of protein structural models using random forests. arXiv:1602.04277
- Cao, R., Wang, Z., Wang, Y., and Cheng, J. (2014). SMOQ: a tool for predicting the absolute residue-specific quality of a single protein model with support vector machines. *BMC Bioinformatics* 15:120. doi: 10.1186/1471-2105-15-120
- Chen, W., Ding, H., Feng, P., Lin, H., and Chou, K. C. (2016). iACP: a sequence-based tool for identifying anticancer peptides. *Oncotarget* 7, 16895–16909. doi: 10.18632/oncotarget.7815
- Chen, W., Feng, P., Yang, H., Ding, H., Lin, H., and Chou, K. C. (2017). iRNA-AI: identifying the adenosine to inosine editing sites in RNA sequences. *Oncotarget* 8, 4208–4217. doi: 10.18632/oncotarget.13758
- Chen, W., Feng, P. M., Lin, H., and Chou, K. C. (2013). iRSpot-PseDNC: identify recombination spots with pseudo dinucleotide composition. *Nucleic Acids Res.* 41:e68. doi: 10.1093/nar/gks1450
- Chen, X.-X., Tang, H., Li, W.-C., Wu, H., Chen, W., Ding, H., et al. (2016). Identification of bacterial cell wall lyases via pseudo amino acid composition. *BioMed Res. Int.* 2016:1654623. doi: 10.1155/2016/1654623
- Chou, K.-C. (2011). Some remarks on protein attribute prediction and pseudo amino acid composition. *J. Theor. Biol.* 273, 236–247. doi: 10.1016/j.jtbi.2010.12.024
- de la Fuente-Nunez, C., Silva, O. N., Lu, T. K., and Franco, O. L. (2017). Antimicrobial peptides: role in human disease and potential as immunotherapies. *Pharmacol. Ther.* 178, 132–140. doi: 10.1016/j.pharmthera.2017.04.002
- Delgado, M., Abad, C., Martinez, C., Leceta, J., and Gomariz, R. P. (2001). Vasoactive intestinal peptide prevents experimental arthritis by downregulating both autoimmune and inflammatory components of the disease. *Nat. Med.* 7, 563–568. doi: 10.1038/87887
- Dhanda, S. K., Gupta, S., Vir, P., and Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clin. Dev. Immunol.* 2013:263952. doi: 10.1155/2013/263952
- Ding, H., Feng, P.-M., Chen, W., and Lin, H. (2014). Identification of bacteriophage virion proteins by the ANOVA feature selection and analysis. *Mol. Biosyst.* 10, 2229–2235. doi: 10.1039/c4mb00316k
- Dreiseitl, S., Ohno-Machado, L., Kittler, H., Vinterbo, S., Billhardt, H., and Binder, M. (2001). A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *J. Biomed. Inform.* 34, 28–36. doi: 10.1006/jbin.2001.1004
- Dubchak, I., Muchnik, I., Holbrook, S. R., and Kim, S. H. (1995). Prediction of protein folding class using global description of amino acid sequence. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8700–8704. doi: 10.1073/pnas.92.19.8700
- Feng, P., Ding, H., Yang, H., Chen, W., Lin, H., and Chou, K. C. (2017). iRNA-PseColl: identifying the occurrence sites of different RNA modifications by incorporating collective effects of nucleotides into PseKNC. *Mol. Ther. Nucleic Acids* 7, 155–163. doi: 10.1016/j.omtn.2017.03.006
- Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., et al. (2017). The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front. Immunol.* 8:278. doi: 10.3389/fimmu.2017.00278
- Gautam, A., Chaudhary, K., Kumar, R., and Raghava, G. P. (2015). Computer-aided virtual screening and designing of cell-penetrating peptides. *Methods Mol. Biol.* 1324, 59–69. doi: 10.1007/978-1-4939-2806-4_4
- Geetha, C., Venkatesh, S. G., Bingle, L., Bingle, C. D., and Gorr, S. U. (2005). Design and validation of anti-inflammatory peptides from human parotid secretory protein. *J. Dent. Res.* 84, 149–153. doi: 10.1177/154405910508400208
- Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1
- Gonzalez, R. R., Fong, T., Belmar, N., Saban, M., Felsen, D., and Te, A. (2005). Modulating bladder neuro-inflammation: RDP58, a novel anti-inflammatory peptide, decreases inflammation and nerve growth factor production in experimental cystitis. *J. Urol.* 173, 630–634. doi: 10.1097/01.ju.0000143192.68223.f7
- Govindaraj, R. G., Manavalan, B., Basith, S., and Choi, S. (2011). Comparative analysis of species-specific ligand recognition in Toll-like receptor 8 signaling: a hypothesis. *PLoS One* 6:e25118. doi: 10.1371/journal.pone.0025118
- Govindaraj, R. G., Manavalan, B., Lee, G., and Choi, S. (2010). Molecular modeling-based evaluation of hTLR10 and identification of potential ligands in Toll-like receptor signaling. *PLoS One* 5:e12713. doi: 10.1371/journal.pone.0012713
- Grieco, P., Rossi, C., Gatti, S., Colombo, G., Carlin, A., Novellino, E., et al. (2005). Design and synthesis of melanocortin peptides with candidacidal and anti-TNF-alpha properties. *J. Med. Chem.* 48, 1384–1388. doi: 10.1021/jm040890j
- Gupta, S., Ansari, H. R., Gautam, A., Open Source Drug Discovery Consortium and Raghava, G. P. (2013). Identification of B-cell epitopes in an antigen for inducing specific class of antibodies. *Biol. Direct* 8:27. doi: 10.1186/1745-6150-8-27
- Gupta, S., Mittal, P., Madhu, M. K., and Sharma, V. K. (2017a). IL17eScan: a tool for the identification of peptides inducing IL-17 response. *Front. Immunol.* 8:1430. doi: 10.3389/fimmu.2017.01430
- Gupta, S., Sharma, A. K., Shastri, V., Madhu, M. K., and Sharma, V. K. (2017b). Prediction of anti-inflammatory proteins/peptides: an insilico approach. *J. Transl. Med.* 15:7. doi: 10.1186/s12967-016-1103-6
- Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747
- Hasan, M. M., Guo, D., and Kurata, H. (2017). Computational identification of protein S-sulfenylation sites by incorporating the multiple sequence features information. *Mol. Biosyst.* 13, 2545–2550. doi: 10.1039/c7mb00491e
- Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003
- Kawashima, S., Pokarowski, P., Pokarowska, M., Kolinski, A., Katayama, T., and Kanehisa, M. (2008). AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.* 36, D202–D205.
- Khan, A., Baharudin, B., Lee, L. H., and Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *J. Adv. Inform. Technol.* 1, 4–20.
- Khondoker, M., Dobson, R., Skirrow, C., Simmons, A., and Stahl, D. (2016). A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. *Stat. Methods Med. Res.* 25, 1804–1823. doi: 10.1177/0962280213502437
- Lai, H. Y., Chen, X. X., Chen, W., Tang, H., and Lin, H. (2017). Sequence-based predictive modeling to identify cancerlectins. *Oncotarget* 8, 28169–28175. doi: 10.18632/oncotarget.15963
- Lata, S., Sharma, B. K., and Raghava, G. P. (2007). Analysis and prediction of antibacterial peptides. *BMC Bioinformatics* 8:263. doi: 10.1186/1471-2105-8-263
- Lee, J., Gross, S. P., and Lee, J. (2013). Improved network community structure improves function prediction. *Sci. Rep.* 3:2197. doi: 10.1038/srep02197
- Lee, J., Lee, K., Joong, I., Joo, K., Brooks, B. R., and Lee, J. (2015). Sigma-RF: prediction of the variability of spatial restraints in template-based modeling by random forest. *BMC Bioinformatics* 16:94. doi: 10.1186/s12859-015-0526-z
- Lin, H., Liang, Z.-Y., Tang, H., and Chen, W. (2017). Identifying sigma70 promoters with novel pseudo nucleotide composition. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2017.2666141 [Epub ahead of print].
- Liu, B., Yang, F., and Chou, K. C. (2017). 2L-piRNA: a two-layer ensemble classifier for identifying piwi-interacting RNAs and their function. *Mol. Ther. Nucleic Acids* 7, 267–277. doi: 10.1016/j.omtn.2017.04.008
- Magnan, C. N., Randall, A., and Baldi, P. (2009). SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* 25, 2200–2207. doi: 10.1093/bioinformatics/btp386

- Manavalan, B., Basith, S., Choi, Y. M., Lee, G., and Choi, S. (2010a). Structure-function relationship of cytoplasmic and nuclear IkappaB proteins: an in silico analysis. *PLoS One* 5:e15782. doi: 10.1371/journal.pone.0015782
- Manavalan, B., Basith, B., Shin, T., Choi, S., Kim, M., and Lee, G. (2017a). MLACP: machine-learning-based prediction of anticancer peptides. *Oncotarget* 8, 77121–77136. doi: 10.18632/oncotarget.20365
- Manavalan, B., Murugapiran, S. K., Lee, G., and Choi, S. (2010b). Molecular modeling of the reductase domain to elucidate the reaction mechanism of reduction of peptidyl thioester into its corresponding alcohol in non-ribosomal peptide synthetases. *BMC Struct. Biol.* 10:1. doi: 10.1186/1472-6807-10-1
- Manavalan, B., Govindaraj, R., Lee, G., and Choi, S. (2011). Molecular modeling-based evaluation of dual function of IkappaBzeta ankyrin repeat domain in toll-like receptor signaling. *J. Mol. Recogn.* 24, 597–607. doi: 10.1002/jmr.1085
- Manavalan, B., Kuwajima, K., Joung, I., and Lee, J. (2015). “Structure-based protein folding type classification and folding rate prediction,” in *Proceedings of the Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on 2015*, Washington, DC.
- Manavalan, B., and Lee, J. (2017). SVMQA: support-vector-machine-based protein single-model quality assessment. *Bioinformatics* 33, 2496–2503. doi: 10.1093/bioinformatics/btx222
- Manavalan, B., Lee, J., and Lee, J. (2014). Random forest-based protein model quality assessment (RFMQA) using structural features and potential energy terms. *PLoS One* 9:e106542. doi: 10.1371/journal.pone.0106542
- Manavalan, B., Shin, T. H., and Lee, G. (2017b). DHSpred: support-vector-machine-based human DNase I hypersensitive sites prediction using the optimal features selected by random forest. *Oncotarget* 9, 1944–1956. doi: 10.18632/oncotarget.23099
- Manavalan, B., Shin, T. H., and Lee, G. (2018). PVP-SVM: sequence-based prediction of phage virion proteins using a support vector machine. *Front. Microbiol.* 9:476. doi: 10.3389/fmicb.2018.00476
- Marie, C., Pitton, C., Fitting, C., and Cavaillon, J. M. (1996). Regulation by anti-inflammatory cytokines (IL-4, IL-10, IL-13, TGFbeta) of interleukin-8 production by LPS- and/ or TNFalpha-activated human polymorphonuclear cells. *Mediators Inflamm.* 5, 334–340. doi: 10.1155/S0962935196000488
- Medzhitov, R. (2008). Origin and physiological roles of inflammation. *Nature* 454, 428–435. doi: 10.1038/nature07201
- Murdoch, J. R., and Lloyd, C. M. (2010). Chronic inflammation and asthma. *Mutat. Res.* 690, 24–39. doi: 10.1016/j.mrfmmm.2009.09.005
- Naggal, G., Usmani, S. S., Dhand, S. K., Kaur, H., Singh, S., Sharma, M., et al. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci. Rep.* 7:42851. doi: 10.1038/srep42851
- Park, K. H., Nan, Y. H., Park, Y., Kim, J. I., Park, I. S., Hahm, K. S., et al. (2009). Cell specificity, anti-inflammatory activity, and plausible bactericidal mechanism of designed Trp-rich model antimicrobial peptides. *Biochim. Biophys. Acta* 1788, 1193–1203. doi: 10.1016/j.bbame.2009.02.020
- Patterson, H., Nibbs, R., McInnes, I., and Siebert, S. (2014). Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. *Clin. Exp. Immunol.* 176, 1–10. doi: 10.1111/cei.12248
- Porto, W. F., Pires, A. S., and Franco, O. L. (2017a). Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnol. Adv.* 35, 337–349. doi: 10.1016/j.biotechadv.2017.02.001
- Porto, W. F., Pires, A. S., and Franco, O. L. (2017b). Antimicrobial activity predictors benchmarking analysis using shuffled and designed synthetic peptides. *J. Theor. Biol.* 426, 96–103. doi: 10.1016/j.jtbi.2017.05.011
- Rayburn, E. R., Ezell, S. J., and Zhang, R. (2009). Anti-inflammatory agents for cancer therapy. *Mol. Cell. Pharmacol.* 1, 29–43. doi: 10.4255/mcpharmacol.09.05
- Saha, I., Maulik, U., Bandyopadhyay, S., and Plewczynski, D. (2012). Fuzzy clustering of physicochemical and biochemical properties of amino acids. *Amino Acids* 43, 583–594. doi: 10.1007/s00726-011-1106-9
- 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. doi: 10.1007/s00253-014-5927-9
- Silva, S., Anuniação, O., and Lotz, M. (2011). “A comparison of machine learning methods for the prediction of breast cancer,” in *Proceedings of the European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics* (Berlin: Springer).
- Steinman, L., Merrill, J. T., McInnes, I. B., and Peakman, M. (2012). Optimization of current and future therapy for autoimmune diseases. *Nat. Med.* 18, 59–65. doi: 10.1038/nm.2625
- Tabas, I., and Glass, C. K. (2013). Anti-inflammatory therapy in chronic disease: challenges and opportunities. *Science* 339, 166–172. doi: 10.1126/science.1230720
- Tang, H., Su, Z. D., Wei, H. H., Chen, W., and Lin, H. (2016). Prediction of cell-penetrating peptides with feature selection techniques. *Biochem. Biophys. Res. Commun.* 477, 150–154. doi: 10.1016/j.bbrc.2016.06.035
- Thakur, N., Qureshi, A., and Kumar, M. (2012). AVPPred: collection and prediction of highly effective antiviral peptides. *Nucleic Acids Res.* 40, W199–W204. doi: 10.1093/nar/gks450
- Wang, X., Yan, R., Li, J., and Song, J. (2016). SOHPRED: a new bioinformatics tool for the characterization and prediction of human S-sulfonylation sites. *Mol. Biosyst.* 12, 2849–2858. doi: 10.1039/c6mb00314a
- Weiner, H. L., Lemere, C. A., Maron, R., Spooner, E. T., Grenfell, T. J., Mori, C., et al. (2000). Nasal administration of amyloid-beta peptide decreases cerebral amyloid burden in a mouse model of Alzheimer's disease. *Ann. Neurol.* 48, 567–579. doi: 10.1002/1531-8249(200010)48:4<567::AID-ANA3>3.0.CO;2-W
- Wu, B. C., Lee, A. H., and Hancock, R. E. W. (2017). Mechanisms of the innate defense regulator peptide-1002 anti-inflammatory activity in a sterile inflammation mouse model. *J. Immunol.* 199, 3592–3603. doi: 10.4049/jimmunol.1700985
- Zhang, Q., Wang, P., Kim, Y., Haste-Andersen, P., Beaver, J., Bourne, P. E., et al. (2008). Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res.* 36, W513–W518. doi: 10.1093/nar/gkn254
- Zhao, Y. W., Su, Z. D., Yang, W., Lin, H., Chen, W., and Tang, H. (2017). IonchanPred 2.0: a tool to predict ion channels and their types. *Int. J. Mol. Sci.* 18:E1838. doi: 10.3390/ijms18091838

Conflict of Interest Statement: 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 interest.

Copyright © 2018 Manavalan, Shin, Kim and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.