ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

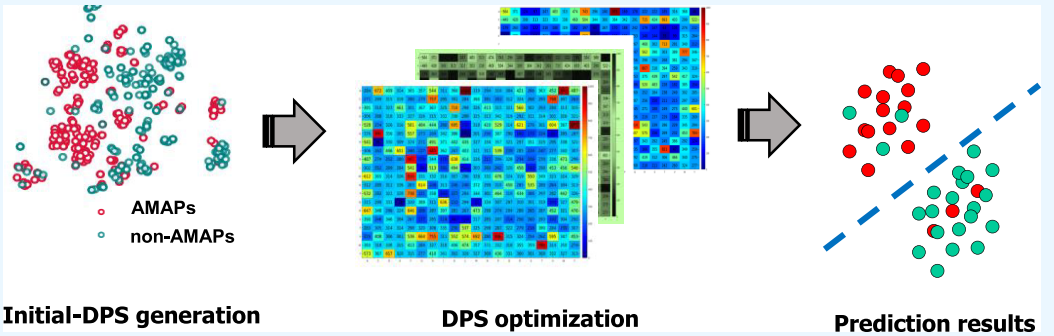
iAMAP-SCM: A Novel Computational Tool for Large-Scale Identification of Antimalarial Peptides Using Estimated Propensity Scores of Dipeptides

Phasit Charoenkwan, Nalini Schaduangrat, Pietro Lio, Mohammad Ali Moni, Pramote Chumnanpuen,* and Watshara Shoombuatong*

Cite This: *ACS Omega* 2022, 7, 41082−41095

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Antimalarial peptides (AMAPs) varying in length, amino acid composition, charge, conformational structure, hydrophobicity, and amphipathicity reflect their diversity in antimalarial mechanisms. Due to the worldwide major health problem concerning antimicrobial resistance, these peptides possess great therapeutic value owing to their low incidences of drug resistance as compared to conventional antibiotics. Although well-known experimental methods are able to precisely determine the antimalarial activity of peptides, these methods are still time-consuming and costly. Thus, machine learning (ML)-based methods that are capable of identifying AMAPs rapidly by using only sequence information would be beneficial for the high-throughput identification of AMAPs. In this study, we propose the first computational model (termed iAMAP-SCM) for the large-scale identification and characterization of peptides with antimalarial activity by using only sequence information. Specifically, we employed an interpretable scoring card method (SCM) to develop iAMAP-SCM and estimate propensities of 20 amino acids and 400 dipeptides to be AMAPs in a supervised manner. Experimental results showed that iAMAP-SCM could achieve a maximum accuracy and Matthew's coefficient correlation of 0.957 and 0.834, respectively, on the independent test dataset. In addition, SCM-derived propensities of 20 amino acids and selected physicochemical properties were used to provide an understanding of the functional mechanisms of AMAPs. Finally, a user-friendly online computational platform of iAMAP-SCM is publicly available at http://pmlabstack.pythonanywhere.com/iAMAP-SCM. The iAMAP-SCM predictor is anticipated to assist experimental scientists in the high-throughput identification of potential AMAP candidates for the treatment of malaria and other clinical applications.

1. INTRODUCTION

Malaria is a global public health concern as it causes millions of deaths annually. Its causative agent, *Plasmodium falciparum*, poses a serious medical challenge.[1] Currently, the development of novel strategies for combating this epidemic disease includes screening of novel antimalarial drugs as a means to curb parasite transmission and thus decrease the mortality rate.[2] Antimicrobial peptides (AMPs) are composed of both naturally and synthetically derived molecules with features such as charge, secondary structure features, molecular weight, and hydrophobicity, to name a few. Even though they are better known for their antimicrobial properties, many AMPs also exhibit broad-spectrum activities toward pathogenic microbiota including the malarial parasite. AMPs target the growth of the malarial parasite at different life cycle stages such as the vertebrate blood stage or the mosquito stage.[3] Polypeptides remain an alternative antimicrobial agent since they have been actively tested against various pathogenic microbes[4] and lead to novel strategies for therapeutic drug development in transgenic malaria-resistant mosquitoes. Besides naturally occurring AMPs and their lineages,

Received: July 15, 2022
Accepted: October 20, 2022
Published: November 2, 2022

significant antimalarial activity has also been demonstrated in peptidomimetic compounds.[5] Antimalarial peptides (AMAPs) consist of several classes of unique peptides i.e., cationic-amphipathic host defense peptides (defensins and cecropins),[6−11] membrane-active peptide antibiotics (gramicidins),[12−15] hydrophobic peptides (cyclosporins),[15,16] thiopeptides (thiostrepton),[17−19] and other natural or synthetic peptides. The aforementioned peptides are involved in membrane interactions that are more specific to the parasite membranes than the host cellular membrane, and some are thought to have more specific intracellular targets.[3] Hence, AMPs play an important role of innate immunity in the host defense mechanism against infections. Most AMPs are encompass cationic molecules at physiological pH and amphipathic peptides that are smaller than 45−50 amino acid residues.[20] In addition, an in vitro study on the antiplasmodial activity of fungal-derived peptide antibiotics has also been reported.[21]

As such, accurate identification of AMAPs will give a direction for elucidating and revealing their potential functional mechanisms. To date, two computational techniques, including molecular docking and molecular dynamics (MD) simulations, have been used in the discovery of new AMAPs or antiplasmodial peptides. Both molecular docking and MD simulations were able to elucidate the interactions between novel phenylalanine−glycine dipeptide sulfonamide conjugate compounds and target protein residues.[22−24] To validate the antiplasmodial activity of cyclic peptides engineered from phytocystatin against falcipain (cysteine proteases from the malarial parasite *P. falciparum*), protein−peptide docking and MD simulations were performed to calculate the free energy of ligand−enzyme binding.[25,26] Moreover, MD experiments of the interaction between a potential drug target PFI1625c (a metalloprotease present in *P. falciparum*) and several bioactive peptides have been performed to screen for AMAPs with high affinity.[27] Although these two computational techniques could enable the discovery of new AMAPs, all these approaches might be limited in the large-scale identification of new AMAPs from a vast number of candidate peptides. Thus, with such potential of AMAPs in therapeutic applications, automated machine learning (ML)-based approaches that can rapidly and accurately identify AMAPs by using only sequence information are highly desirable.

In this study, we present a new computational approach (named iAMAP-SCM) that is able to rapidly identify AMAPs and automatically generate propensities of 20 amino acids and 400 dipeptides to be AMAPs. To the best of our knowledge, iAMAP-SCM is the first scoring card method (SCM)-based model for AMAP identification and characterization. In iAMAP-SCM, we established a benchmark dataset by collecting experimentally validated AMAPs and non-AMAPs extracted from several resources. Then, propensities of 20 amino acids and 400 dipeptides were estimated and optimized using the 10-fold cross-validation scheme. After that, the estimated propensities of 20 amino acids were utilized to select important physicochemical properties (PCPs) from the amino acid index database (AAindex). Finally, the optimal propensities of 400 dipeptides (DPS) were selected and used to develop the final predictive model (iAMAP-SCM). Experimental results on the independent test dataset showed that iAMAP-SCM could achieve a maximum accuracy (ACC) and Matthew's coefficient correlation(MCC) of 0.957 and 0.834, respectively. In addition, comparative analysis results revealed that iAMAP-SCM outperformed decision tree (DT)-based, k-nearest neighbor (KNN)-based, and logistic regression (LR)-based classifiers as indicated by MCC and balanced accuracy (BACC) on both the training and independent test datasets. Furthermore, analysis results based on the 20 amino acid propensities and selected PCPs indicated four important characteristics of AMAPs, which can be encapsulated as follows: (i) AMAPs have high conformational and thermal stability; (ii) AMAPs are composed of hydrophobic amino acids; (iii) AMAPs are composed of amino acids interacting with the cell membrane; and (iv) AMAPs tend to be in the amphipathic antimicrobial peptide class.

2. MATERIALS AND METHODS

2.1. Dataset Preparation. Dataset preparation is an important step in the development of ML-based models. Herein, we created two datasets called main and alternative datasets. To be specific, positive samples were experimentally validated AMAPs, which were mainly derived from ParaPep[28] and several literature studies.[10,29−33] After removing redundant samples, a total of 139 unique AMAPs were accounted for in this study. These AMAPs were then used for creating the main and alternative datasets. Since there are very few experimentally validated non-AMAPs, negative samples in the main dataset were derived from non-parasitic peptides, which were derived from the work of Xiao et al.[34] For the alterative dataset, negative samples were derived from random peptides in SwissProt, which were derived from the work of Agrawal et al.[35] Herein, we excluded the negative samples found in both main and alterative datasets. As a result, the main dataset consisted of 139 positives and 2135 negatives, while the alternative dataset consisted of 139 positives and 677 negatives.

2.2. Construction of Training and Independent Test Datasets. Herein, around 80% of positive and negative samples were used for constructing the training dataset, while the remaining were considered for constructing the independent test dataset.

In general, the training dataset is employed for model construction and optimization, while the independent test dataset is employed to evaluate the effectiveness and generalizability of the model. Therefore, for the main dataset, its training and independent test datasets (referred herein as Main-CV and Main-TS, respectively) contained (111 positives and 1708 negatives) and (28 positives and 427 negatives), respectively. In the same way, the training and independent test datasets of the alternative dataset (referred herein as Alter-CV and Alter-TS, respectively) contained (111 positives and 542 negatives) and (28 positives and 135 negatives), respectively. Table S1 summarizes the details of the Main-CV, Main-TS, Alter-CV, and Alter-TS datasets. Herein, we noticed that most of the AMAPs contain less than 30 amino acid residues (87.27%), while most of the non-AMAPs contain less than 40 amino acid residues (72.19 and 73.62% for the Main-CV and Alter-CV datasets, respectively) (Tables S2 and S3).

2.3. General Framework of iAMAP-SCM. Here, iAMAP-SCM is an SCM-based predictor coupled with the optimal propensities of 400 dipeptides designed for AMAP prediction and characterization. Numerous previous studies demonstrated that the SCM method iAMAP-SCM is able to rapidly identify several biological functions of peptides and automatically generate propensities of 20 amino acids and 400

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

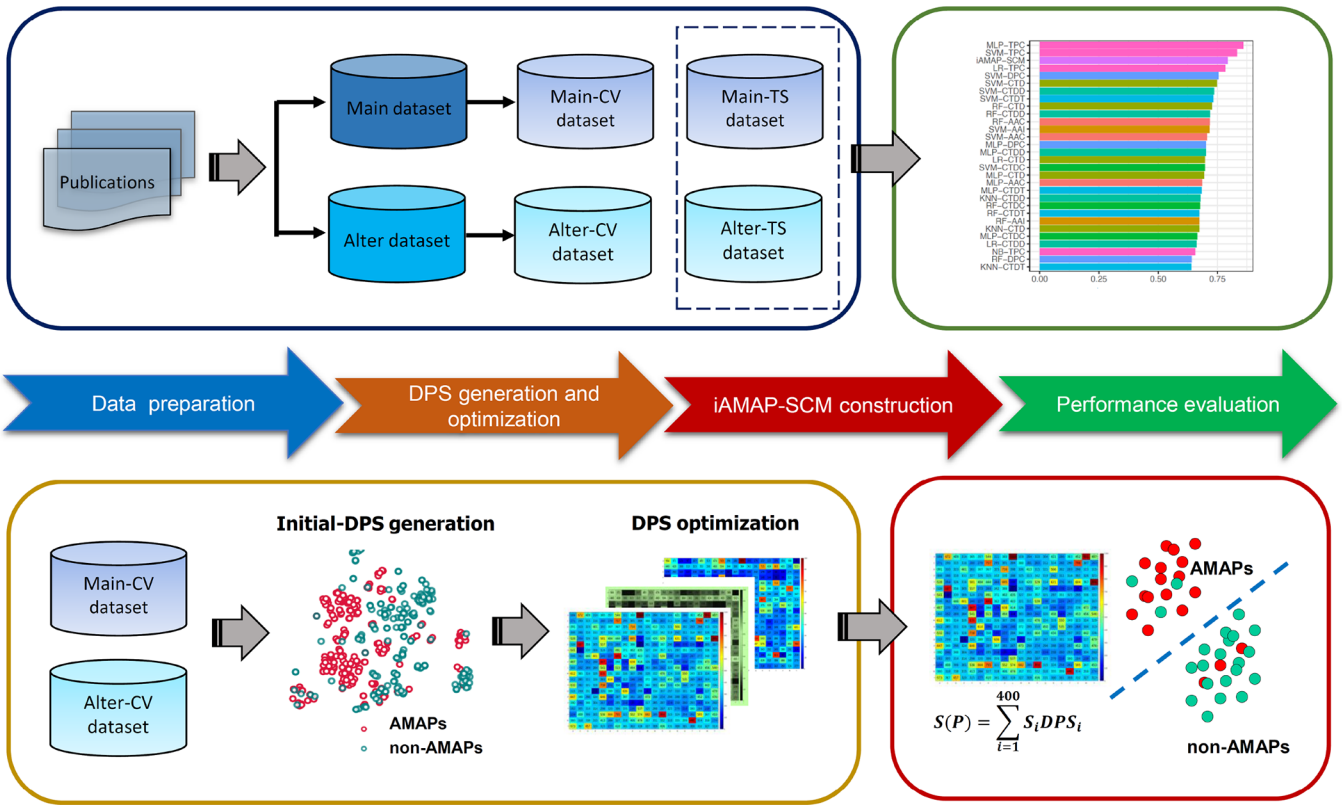


Figure 1. Schematic framework of the development of iAMAP-SCM. The schematic framework of the development of SCMB3PP contains four main steps: (i) training and independent dataset preparation, (ii) DPS generation and optimization, (iii) iAMAP-SCM construction, and (iv) performance evaluation.

dipeptides as a function of interest in a supervised manner.[36−41] The iAMAP-SCM construction involves three main steps (as summarized in Figure 1), which involves the generation of an initial propensity score of 400 dipeptides (init − DPS), DPS optimization, iAMAP-SCM construction, and AMAP characterization using informative PCPs from the AAindex database. Below, we briefly describe the calculation and optimization procedures of the DPS. In the meanwhile, the propensities of 400 *g*-gap dipeptides (GDPS) ($g = i$, where i = 1, 2, 3, 4, and 5) can be optimized in the same procedure as the DPS.

2.3.1. Generating Init-DPS. We generated an initial-DPS matrix (init − $DPS_{ij} = \{n_{ij}\}$) using a simple statistical method, where n_{ij} is the number of occurrences of the dipeptides containing the i^{th} amino acid followed by the j^{th} amino acid, where i, j = 1,2,3, ...,20. The first step is to calculate the number of occurrences of 400 dipeptides in positive ($P_{ij} = (n_{ij} \mid C = 1)$) and negative ($N_{ij} = (n_{ij} \mid C = 0)$) samples. The second step is to compute the composition of P_{ij} and N_{ij}, which are referred to as cP_{ij} and cN_{ij}. The third step is to compute the scores of 400 dipeptides by subtracting cP_{ij} from cN_{ij}. Finally, the scores of 400 dipeptides are normalized into the range of 0−1000 to construct the init − DPS, which is a 20 × 20 matrix containing initial propensity scores of 400 dipeptides.

2.3.2. Optimizing DPS. Since the information of init − DPS might be not sufficient to enable the construction of a comprehensive predictive model,[38,42,43] we employed the genetic algorithm (GA) to create the DPS matrix, which is a 20 × 20 matrix containing optimal propensity scores of 400 dipeptides. The major contribution of the DPS over init − DPS could be listed as follows: (i) the DPS could enhance the discriminative ability and (ii) the DPS could conserve the information of AMAPs. In this work, the fitness function of the GA algorithm is represented by

$$\mathrm{Max(DPS)} = 0.9 \times \mathrm{AUC} + 0.1 \times R \quad (1)$$

As can be seen, this fitness function involves the area under the receiver operating characteristics (ROC) curve (AUC) and Pearson's correlation coefficient (R-value) between DPS and init − DPS.[36,38] To control overfitting and biasness, we assessed the predictive performance of each DPS by the 10-fold cross-validation scheme. More details of the DPS optimization is summarized in the Supplementary Methods.

2.3.3. Constructing iAMAP-SCM. In this step, we aimed to construct a scoring function $S(P)$ coupled with the DPS and to determine an optimal threshold value. As a result, the scoring function $S(P)$ used for AMAP identification is defined as follows:

$$S(P) = \sum_{i=1}^{400} S_i \mathrm{DPS}_i \quad (2)$$

where S_i and DPS_i represent the occurrence number and the propensity score of the i^{th} dipeptide. The query peptide P is considered as AMAP, if $S(P)$ is greater than the threshold value, otherwise the query peptide P is considered as non-AMAP

$$\begin{cases} \text{AMAP}, S(P) > \text{threshold value} \\ \text{non-AMAP}, S(P) < \text{threshold value} \end{cases} \quad (3)$$

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

The DPS having the highest MCC were considered for constructing the final model (iAMAP-SCM). More detailed information for the SCM method and tDPS optimization is recorded in our previous studies.[36,38,42,43]

2.3.4. AMAP Characterization Using Informative PCPs. After obtaining the optimal DPS, the propensity score of the i^{th} amino acid is computed by averaging the propensity scores of all dipeptides containing the i^{th} amino acid. These propensities were then utilized to select informative PCPs relating to functional mechanisms of AMAPs. To select informative PCPs, first, we collected 544 PCPs from the AAindex database[44] and the PCPs having a value of "NA" were eliminated to optimize the utilization of PCPs. As a result, the remaining 531 PCPs were employed herein. Second, we calculated the *R*-value between the propensities of the 20 amino acids and each of the 531PCPs. Finally, we selected the top 20 informative PCPs with the highest *R*-values for AMAP characterization.

2.4. Conventional ML-Based Classifiers. In this study, a total of 72 different ML classifiers (8 MLs × 9 descriptors) were created by using eight popular ML algorithms (DT, KNN, LR, multilayer perceptron (MLP), naive Bayes (NB), partial least squares regression (PLS), random forest (RF), and support vector machine (SVM)) trained with nine conventional feature descriptors (AAindex, PCP, amino acid composition (AAC), composition transition and distribution (CTD), composition (CTDC), distribution (CTDD), transition (CTDT), dipeptide composition (DPC), and tripeptide composition (TPC)). For implementation, all the 72 ML classifiers were trained and optimized under the 10-fold cross-validation test based on the iFeature[45] and the Scikit-learn v0.22.0 packages[46] (Table S4). A detailed description of the construction of these ML classifiers is recorded in our previous reports.[42,43,47,48]

2.5. Evaluation Metrics. The performance of iAMAP-SCM and conventional ML classifiers was evaluated in terms of five well-known performance measures, including ACC, AUC, BACC, MCC, sensitivity (Sn), specificity (Sp), and area under the precision-recall (PR) curve (AUPR),[47,49] described as follows:

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \quad (4)$$

$$\mathrm{Sn} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \quad (5)$$

$$\mathrm{Sp} = \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})} \quad (6)$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} \quad (7)$$

$$\mathrm{BACC} = 0.5 \times \mathrm{Sn} + 0.5 \times \mathrm{Sp} \quad (8)$$

where TPs and TNs represent the number of true AMAPs and true non-AMAPs that are correctly predicted as AMAPs and non-AMAPs, respectively. On the other hand, FPs and FNs represent the number of true non-AMAPs and true AMAPs that are predicted as AMAPs and true non-AMAPs, respectively.[50−54]

3. RESULTS AND DISCUSSION

3.1. Performance Evaluation of BLAST-Based Predictors. In this section, we created a BLAST-based predictor for the prediction of AMAPs. A detailed construction of the BLAST-based predictor is summarized in our previous studies.[55,56] To assess the performance of BLAST-based predictors, the expectation values (*E*-values) in the range of 0.0001 to 0.1 were employed. Table 1 provides the performance comparison results of various *E*-values on the Main-TS and Alter-TS datasets. The highest ACC values of 0.901 and 0.889 on the Main-TS and Alter-TS datasets were achieved using *E*-value cutoff values of 0.0001 and 0.1, respectively. However, these cutoff values were found to have Sn values of 0.222 and 0.407 on the Main-TS and Alter-TS datasets, respectively, thereby indicating that the BLAST-based predictor was not effective for accurately identifying true AMAPs. Altogether, ML-based predictors that are capable of accurately and rapidly identifying AMAPs are more suitable for this problem.

Table 1. Performance Evaluation of BLAST-Based Predictors Using Various *E*-Value Cutoffs

dataset	*E*-value	ACC	BACC	Sn	Sp	MCC
Main-TS	0.1	0.824	0.629	0.407	0.850	0.164
	0.01	0.872	0.602	0.296	0.909	0.159
	0.001	0.892	0.596	0.259	0.932	0.167
	0.0001	0.901	0.583	0.222	0.944	0.158
Alter-TS	0.1	0.889	0.696	0.407	0.985	0.539
	0.01	0.883	0.648	0.296	1.000	0.510
	0.001	0.877	0.630	0.259	1.000	0.475
	0.0001	0.870	0.611	0.222	1.000	0.439

3.2. Performance Evaluation of Different Sets of Propensity Scores. Here, we evaluated and analyzed the predictive capability of variant SCM classifiers coupled with six different types of propensity scores by performing 10-fold cross-validation and independent tests on both the main and alternative datasets. Tables 2 and 3 list a comparison of the predictive performance of the six different types of propensity scores. Table 2 shows that the highest MCC of 0.793 in terms of the Main-CV dataset is achieved by using DPS, while the second and third highest MCC of 0.765 are obtained from GDPS (g = 1) and GDPS (g = 4). Interestingly, we noticed that the best-performing DPS also accomplished satisfactory results with the highest MCC of 0.776 on the Main-TS dataset. For the performance on the alternative dataset, we noticed that DPS performed well with the highest MCC of 0.917 on the Alter-CV dataset (Table 3), while GDPS (g = 1) performed well with the third highest MCC of 0.900. Remarkably, DPS still yielded an impressive predictive performance with an MCC of 0.834 on the Alter-TS dataset. Altogether, our comparative results highlighted the importance and contribution of DPS to the performance of the SCM classifier. Thus, for the convenience of discussion, the SCM classifiers combined with DPS is referred to as iAMAP-SCM. The propensities of 400 dipeptides based on threshold values of 449 and 309 on the Main-CV and Alter-CV datasets, respectively, are shown in Figure 2.

As mentioned in the General Framework of iAMAP-SCM Section, the GA method was used to optimize the initial propensity scores as means to improve the performance of the SCM classifier. To elucidate this important point, we

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

Table 2. Performance Comparison of Different Types of Propensity Scores on the Main Dataset

dataset	descriptor	cutoff	ACC	BACC	Sn	Sp	MCC	AUC
Main-CV	DPS	449	0.978	0.857	0.720	0.995	0.793	0.881
	GDPS (g = 1)	424	0.978	0.852	0.710	0.994	0.765	0.924
	GDPS (g = 2)	540	0.978	0.840	0.687	0.994	0.760	0.911
	GDPS (g = 3)	513	0.980	0.820	0.643	0.997	0.759	0.899
	GDPS (g = 4)	478	0.980	0.814	0.631	0.998	0.765	0.897
	GDPS (g = 5)	586	0.979	0.783	0.568	0.998	0.709	0.854
Main-TS	DPS	449	0.978	0.826	0.654	0.998	0.776	0.820
	GDPS (g = 1)	424	0.967	0.775	0.560	0.991	0.644	0.856
	GDPS (g = 2)	540	0.971	0.769	0.545	0.993	0.647	0.798
	GDPS (g = 3)	513	0.982	0.824	0.650	0.998	0.769	0.865
	GDPS (g = 4)	478	0.973	0.795	0.600	0.991	0.657	0.850
	GDPS (g = 5)	586	0.980	0.777	0.556	0.998	0.702	0.819

Table 3. Performance Comparison of Different Types of Propensity Scores on the Alternative Dataset

dataset	descriptor	cutoff	ACC	BACC	Sn	Sp	MCC	AUC
Alter-CV	DPS	309	0.977	0.950	0.910	0.991	0.917	0.968
	GDPS (g = 1)	423	0.975	0.935	0.877	0.993	0.900	0.957
	GDPS (g = 2)	390	0.972	0.925	0.860	0.991	0.886	0.943
	GDPS (g = 3)	312	0.979	0.940	0.885	0.994	0.912	0.971
	GDPS (g = 4)	329	0.968	0.915	0.840	0.989	0.862	0.962
	GDPS (g = 5)	418	0.958	0.884	0.784	0.983	0.806	0.935
Alter-TS	DPS	309	0.957	0.896	0.808	0.985	0.834	0.903
	GDPS (g = 1)	423	0.956	0.876	0.760	0.993	0.826	0.873
	GDPS (g = 2)	390	0.943	0.834	0.682	0.985	0.745	0.799
	GDPS (g = 3)	312	0.974	0.900	0.800	1.000	0.881	0.948
	GDPS (g = 4)	329	0.942	0.839	0.700	0.978	0.727	0.911
	GDPS (g = 5)	418	0.967	0.885	0.778	0.993	0.835	0.941

compared the predictive performance of DPS to that of initial-DPS (Table 4). It could be noticed that DPS yielded the overall best performance as compared with initial-DPS as judged by all measures on both the 10-fold cross-validation and independent tests. To be specific, the BACC and MCC of DPS were 2.2−6.0 and 7.0−17.3% higher than those of the initial-DPS, respectively, on the Main-TS and Alter-TS datasets. Altogether, our analysis results confirmed that the propensities of 400 dipeptides (DPS) provide more discriminative power for identifying AMAPs than initial-DPS.

3.3. Comparison of iAMAP-SCM with Conventional ML Classifiers. Since iAMAP-SCM represents the first computational approach designed for AMAP identification, we compared its performance to that of other conventional ML classifiers as means to elucidate its predictive capability and effectiveness. Herein, the conventional ML classifiers, which contained a total 72 ML classifiers trained with eight ML algorithms and nine features were created. The performance evaluation results of the 72 ML classifiers are recorded in Tables S5−S8. Furthermore, to simplify the comparative analysis, we compared the performance of iAMAP-SCM to that of five top-ranked ML classifiers as determined by MCC values on the training datasets. These comparative analysis results are provided in Figures 3−5 and Table 5.

Figure 3 shows that the five top-ranked ML classifiers on the Main-CV dataset contain MLP-TPC, SVM-TPC, LR-TPC, SVM-DPC, and SVM-CTD, while the five top-ranked ML classifiers on the Alter-CV dataset contain MLP-TPC, LR-DPC, SVM-TPC, SVM-CTD, and SVM-AAC. This demonstrates that MLP-TPC is the most powerful ML combination for AMAP prediction in terms of both the Main-CV and Alter-CV datasets. From Figure 3, iAMAP-SCM is found to yield the third and first highest MCC of 0.793 and 0.917 for the Main-CV and Alter-CV datasets, respectively. In the meanwhile, the highest MCC in terms of the Main-CV was found in MLP-TPC. For the Main-TS and Alter-TS datasets, iAMAP-SCM and MLP-TPC achieved a similar predictive performance and outperformed the remaining top-ranked ML classifiers as indicated by BACC and MCC.

Although our iAMAP-SCM approach provided a lower AUPR value as compared with MLP-TPC, SVM-DPC, and SVM-CTD (0.737 versus 0.787−0.794 and 0.869 versus 0.867−0.928 for the Main-TS and Alter-TS datasets, respectively), the major advantages of iAMAP-SCM over the compared ML classifiers could be summarized in three aspects: (i) iAMAP-SCM achieved a competitive performance on the Main-TS and Alter-TS datasets compared with MLP-TPC. In the meanwhile, iAMAP-SCM outperformed LR, KNN, and DT classifiers in terms of MCC for the Main-TS and Alter-TS datasets; (ii) iAMAP-SCM is a simple and easy-to-understand approach for biologists and biochemists, and (iii) unlike MLP and SVM methods that are known as black-box computational approaches, iAMAP-SCM is capable of estimating propensities of 20 amino acids and 400 dipeptides. This information is beneficial for AMAP characterization.

3.4. Case Studies. In this section, we further utilized our iAMAP-SCM approach to perform case studies as a means to verify its predictive capability in practical situations. First, we collected experimentally validated AMAPs from various literature studies.[57−60] Second, to alleviate the over-estimation and over-fitting issues, these AMAPs found in the main and alternative datasets were excluded. As a result, the remaining

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

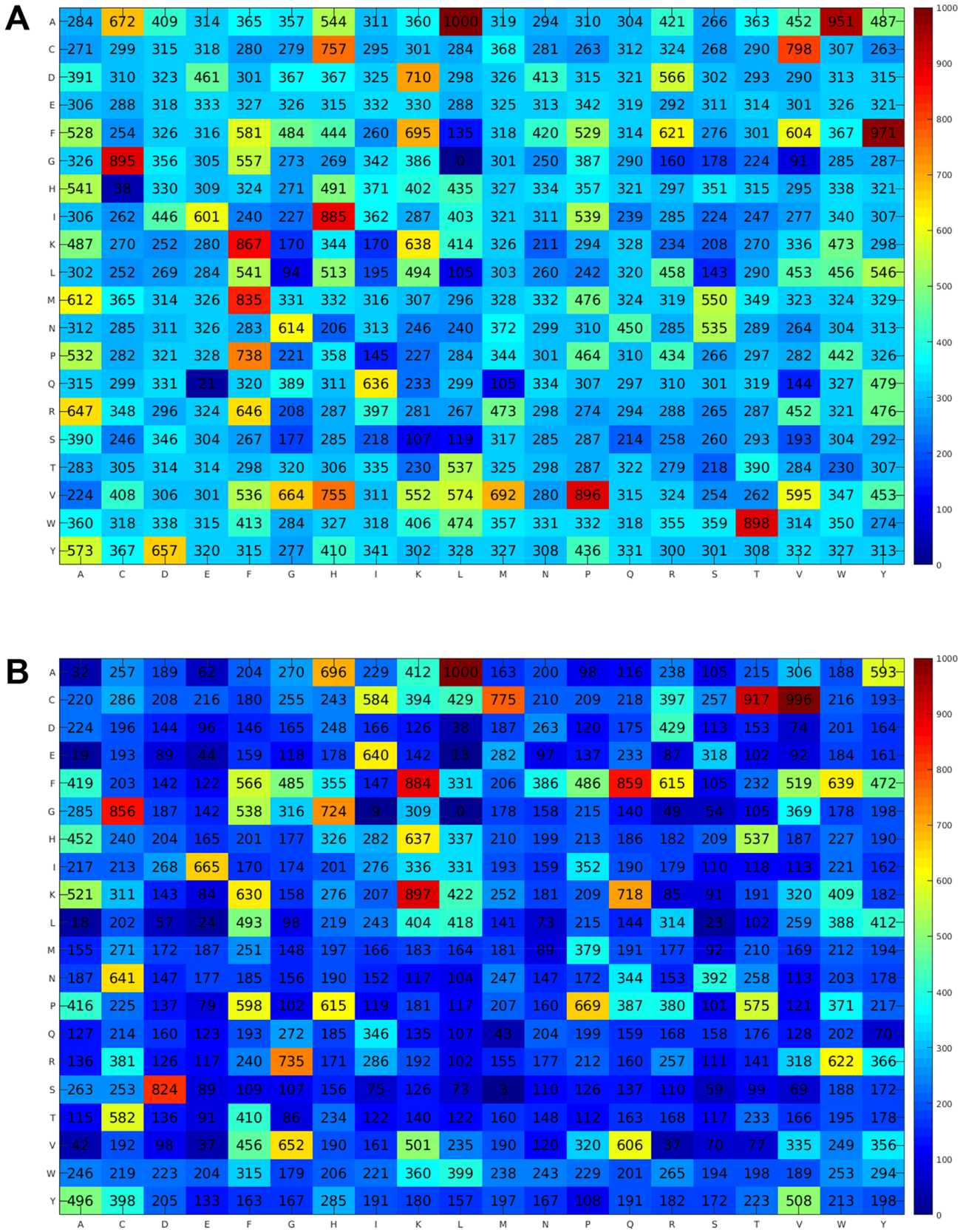


Figure 2. Propensity scores of 20 amino acids to be AMAPs obtained from the proposed iAMAP-SCM based on the Main-CV (A) and Alter-CV (B) datasets.

32 AMAPs were obtained and accounted for the case studies (as listed in Table S9). As mentioned above, iAMAP-SCM was optimized and constructed by using the Main-CV and Alter-CV datasets. Therefore, iAMAP-SCM constructed by these two datasets was performed on the case studies separately. From Tables S10 and S11, several observations

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

Table 4. Performance Comparison of DPS with Initial-DPS on the Main and Alternative Datasets

dataset	descriptor	ACC	BACC	Sn	Sp	MCC	AUC
Main-CV	Initial-DPS	0.958	0.757	0.530	0.985	0.588	0.895
	DPS	0.978	0.857	0.720	0.995	0.793	0.881
Main-TS	Initial-DPS	0.971	0.804	0.615	0.993	0.706	0.776
	DPS	0.978	0.826	0.654	0.998	0.776	0.820
Alter-CV	Initial-DPS	0.940	0.885	0.801	0.969	0.786	0.953
	DPS	0.977	0.950	0.910	0.991	0.917	0.968
Alter-TS	Initial-DPS	0.907	0.836	0.731	0.941	0.661	0.858
	DPS	0.957	0.896	0.808	0.985	0.834	0.903

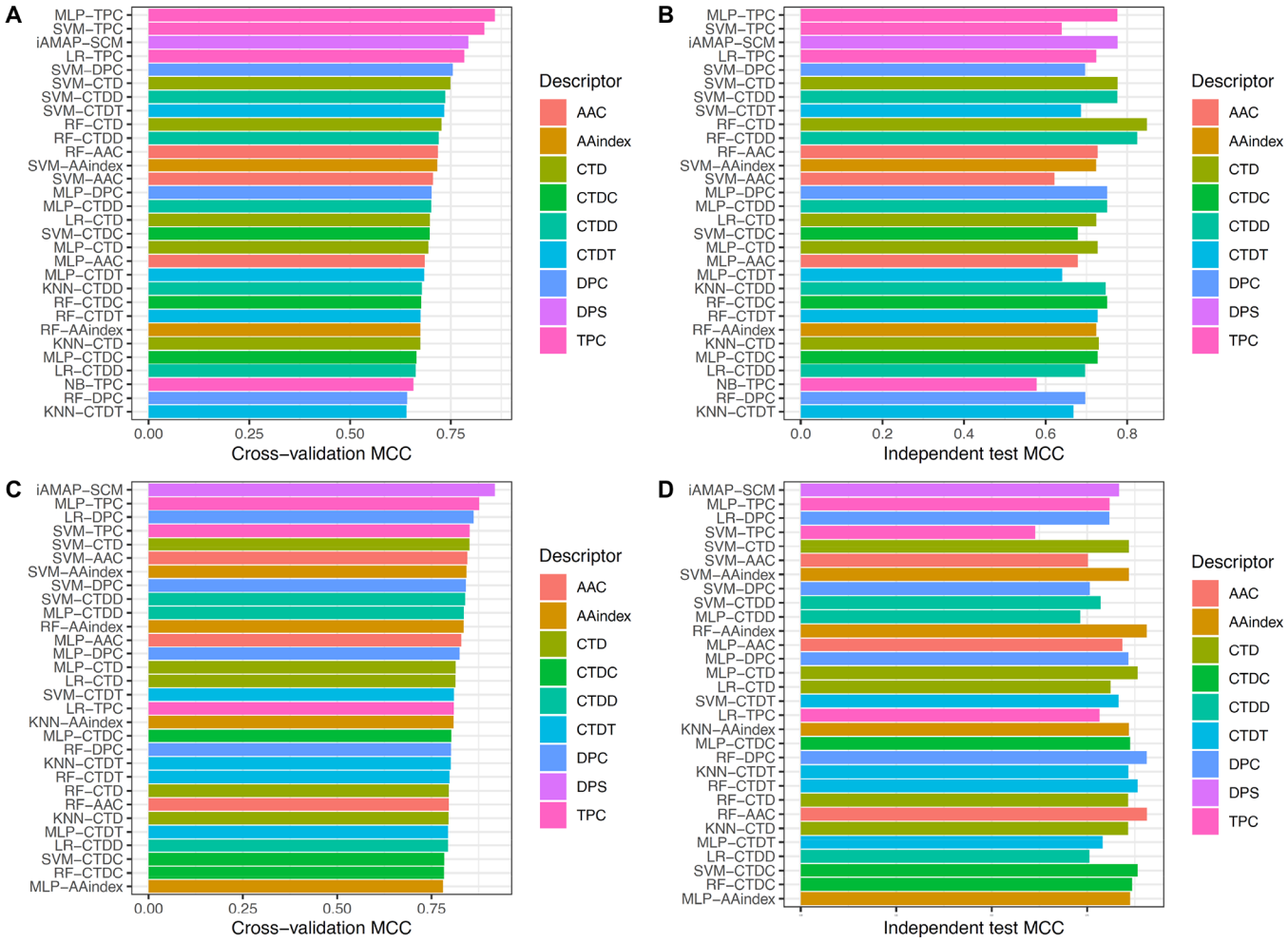


Figure 3. Performance comparison between iAMAP-SCM and top 20 ML classifiers. Comparisons of MCC values on the Main-CV (A), Main-TS (B), Alter-CV (C), and Alter-TS (D) datasets.

can be made: (i) 20 and 31 samples (ACC of 0.625 and 0.969, respectively) are correctly identified as AMAPs by iAMAP-SCM developed by using the Main-CV and Alter-CV datasets, respectively; (ii) MLP-TPC developed by using the Main-CV and Alter-CV datasets can correctly identify 21 and 24 samples (ACC of 0.656 and 0.750, respectively), respectively; (iii) Based on the case studies, we noticed that the performance of the predictive models (i.e., MLP-TPC, SVM-CTD, and iAMAP-SCM) trained using the Alter-CV dataset is better than that of the predictive models trained using the Main-CV dataset, demonstrating that the Alter-CV dataset could be more suitable for creating an efficient predictive model. In addition, this result indicates that iAMAP-SCM outperformed the compared methods and could effectively determine candidates as AMAPs from large-scale proteins.

3.5. Characterization of Antimalarial Peptides Using SCM-Derived Propensity Scores. In this section, we employed SCM-derived amino acid propensities for AMAP characterization. As mentioned above, the SCM-derived amino acid propensities were generated using the optimal propensities of 400 dipeptides to be AMAPs (Figure 2A) and a straightforward statistical approach.[38,42,43] As can be seen from Table 6, the top five informative amino acids for AMAPs are Phe, Ala, Val, His, and Trp. Interestingly, the unique non-polar amino acids (Phe, Ala, Trp, and Val) and positive charge amino acid (His) are prevalent in AMAPs as compared to non-AMAPs. Similar to the amino acid composition results

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

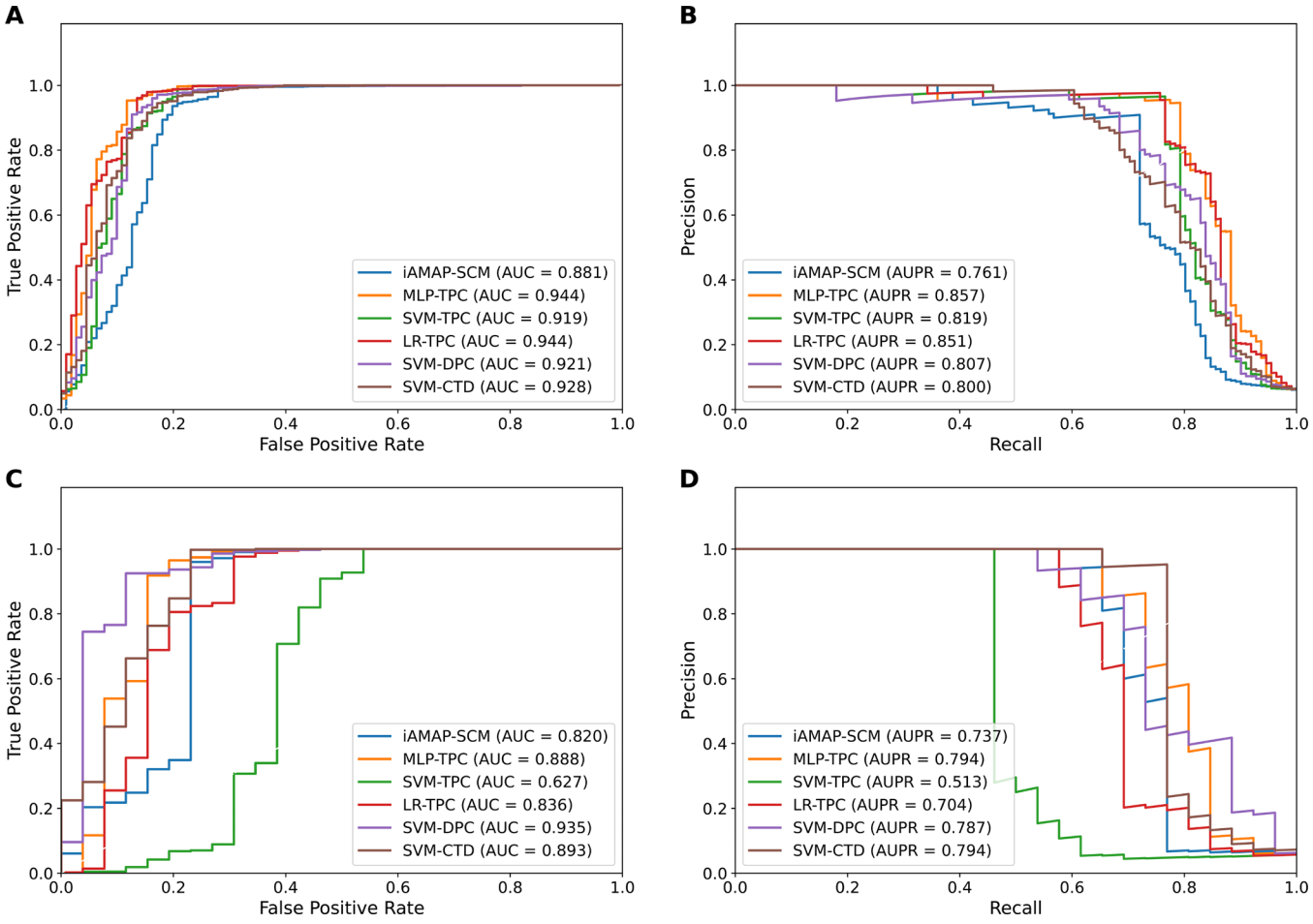


Figure 4. Performance comparison between iAMAP-SCM and top five ML classifiers. Comparisons of the ROC curve, AUC value, PR curve, and AUPR value on the Main-CV (A, B) and Main-TS (C, D) datasets.

that have been reported before, the preferential amino acids that are more prominent in AMPs and AMAPs are Ala, Lys, Arg, Ile, Leu, Phe, Pro, Tyr, and Val.[31,33,61,62]

For the amphipathic feature of AMPs, peptide interactions with the membranes are driven by both the hydrophilic and lipophilic nature of the peptides. The positive charge or cationic side chain amino acids (Arg, Lys, and His) and bulky (Pro, Phe, Val, Ile and Trp) side chains are fundamental for peptide antimicrobial activity.[63] This is also correlated to several pieces of evidence that have reported that the amino acids Ala, Lys, Arg, Ile, Leu, Phe, Pro, Tyr, and Val are frequently included on antimicrobial and antimalarial peptides. In the case of the hydrophobic character of AMPs, Val, Phe, and Pro are considered to be important for collapsing the plasmodium cell membrane.[62,64−66] Research observed that deletion of Val could lead to a 23% decrease in the antiplasmodial activity.[62] Since Phe is necessary for proficient antibacterial activity and membrane phospholipid interaction,[66] the substitution of polar residues (Arg or Lys) by Phe could promisingly increase antimalarial activity.[67] Furthermore, a study observed the parasite specific property of Phe in combination with Pro to show lower toxicity toward healthy cells such as erythrocytes and fibroblasts.[68] For the positively charged amino acids with the fourth highest propensity score (His), it was observed that His-rich amphipathic cationic peptides could significantly enhance membrane disruption and antimicrobial activities including their antimalarial ability.[68]

3.6. Characterization of Antimalarial Peptides Using Informative Physicochemical Properties. In this section, iAMAP-SCM coupled with the SCM-derived amino acid propensities was applied to select informative PCPs for AMAP characterization. Table S6 lists the top 20 PCPs with the highest *R*-values, which are perceived as most important for AMAP characterization. As seen in Table S6, the top five PCPs with the highest *R*-values consisted of TAKK010101, BIOV880101, RICJ880111, BIOV880102, and CIDH920102, respectively. Detailed analysis results of these important PCPs for AMAPs are provided in the following subsections.

3.6.1. AMAPs Have High Conformational and Thermal Stability. The TAKK010101 can be described as "side-chain contribution to protein stability (kJ/mol)".[69] The stability estimation profile is a powerful tool that has shown to be useful in the analysis of protein sequences. Since the side-chain contribution to protein stability includes the entropic effect, they might represent the contributions of each amino acid side chain to the conformational stability of the globular protein. However, the high *R*-value implies that hydrophobic amino acids contribute largely to AMAPs. Notably, the five top-ranked amino acids were Trp, Phe, Ile, Tyr, and Leu, possessing the highest side chain hydrophobicity values.[70] As previously mentioned above, from the amino acid propensity results, the overall stability of AMAPs is high, with the five top-ranked amino acids being highly hydrophobic (non-polar) side chains. Research has suggested that Trp and Phe are

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

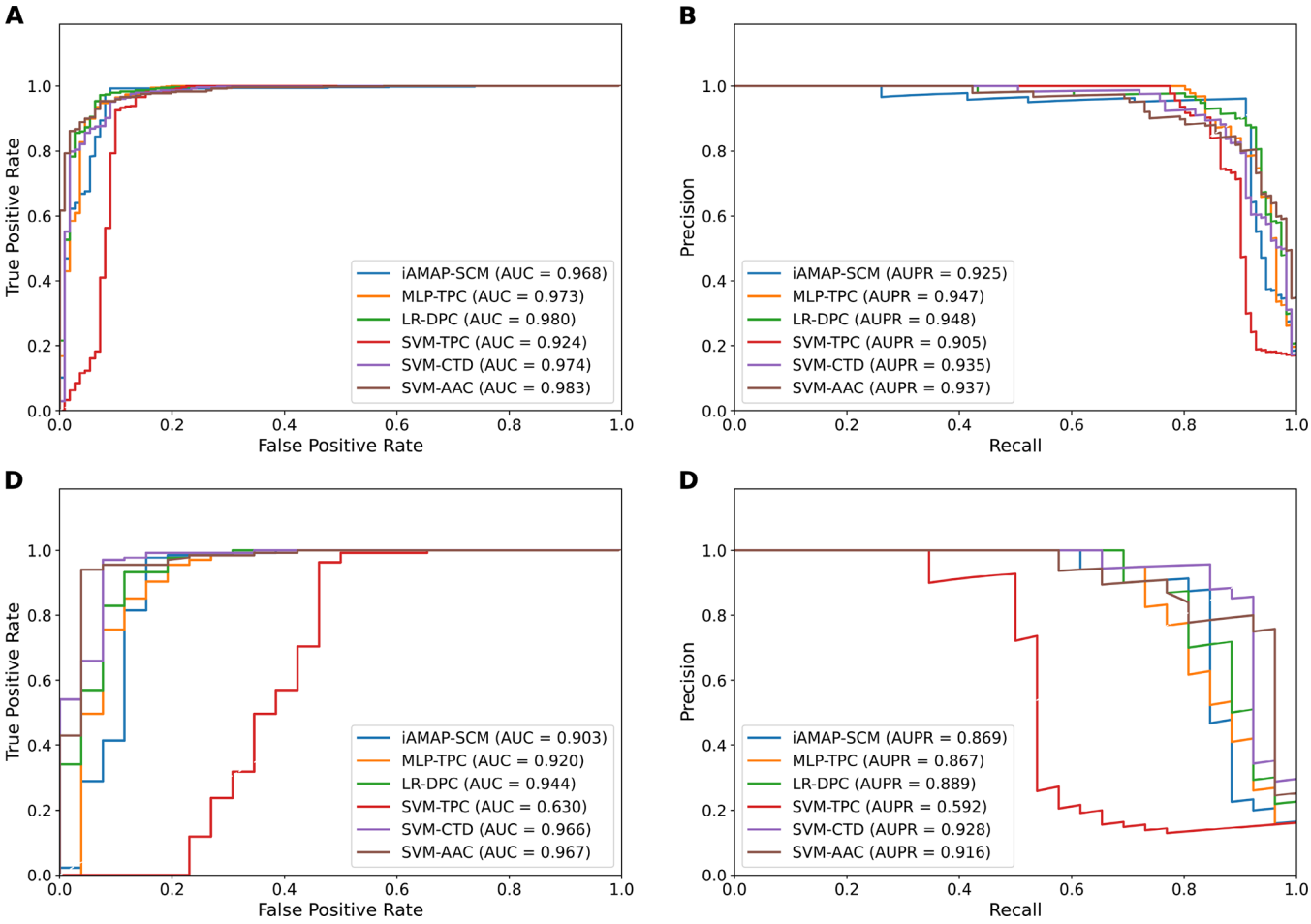


Figure 5. Performance comparison between iAMAP-SCM and top five ML classifiers. Comparisons of the ROC curve, AUC value, PR curve, and AUPR value on the Alter-CV (A, B) and Alter-TS (C, D) datasets.

Table 5. Performance Comparison of iAMAP-SCM and Best-Performing ML Classifiers on the Main and Alternative Datasets

dataset	method	ACC	BACC	Sn	Sp	MCC	AUC
Main-CV	MLP-TPC	0.985	0.894	0.792	0.997	0.859	0.944
	SVM-TPC	0.982	0.868	0.737	0.998	0.833	0.919
	LR-TPC	0.976	0.881	0.773	0.989	0.783	0.944
	SVM-DPC	0.975	0.805	0.611	0.999	0.754	0.921
	SVM-CTD	0.974	0.817	0.638	0.996	0.749	0.928
	iAMAP-SCM	0.978	0.857	0.720	0.995	0.793	0.881
Main-TS	MLP-TPC	0.978	0.808	0.615	1.000	0.775	0.888
	SVM-TPC	0.967	0.712	0.423	1.000	0.639	0.627
	LR-TPC	0.974	0.787	0.577	0.998	0.724	0.836
	SVM-DPC	0.971	0.750	0.500	1.000	0.697	0.935
	SVM-CTD	0.978	0.826	0.654	0.998	0.776	0.893
	iAMAP-SCM	0.978	0.826	0.654	0.998	0.776	0.820
Alter-CV	MLP-TPC	0.966	0.915	0.837	0.993	0.876	0.973
	LR-DPC	0.962	0.912	0.837	0.987	0.861	0.980
	SVM-TPC	0.960	0.886	0.773	0.998	0.851	0.924
	SVM-CTD	0.959	0.903	0.818	0.987	0.850	0.974
	SVM-AAC	0.954	0.929	0.891	0.967	0.844	0.983
	iAMAP-SCM	0.977	0.950	0.910	0.991	0.917	0.968
Alter-TS	MLP-TPC	0.950	0.846	0.692	1.000	0.808	0.920
	LR-DPC	0.950	0.877	0.769	0.985	0.808	0.944
	SVM-TPC	0.907	0.727	0.462	0.993	0.613	0.630
	SVM-CTD	0.963	0.916	0.846	0.985	0.859	0.966
	SVM-AAC	0.932	0.882	0.808	0.956	0.752	0.967
	iAMAP-SCM	0.957	0.896	0.808	0.985	0.834	0.903

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

Table 6. Propensity Scores of 20 Amino Acids to be AMAPs (PS) along with Different Amino Acid Compositions (%) of AMAPs and Non-AMAP Based on the Main-CV Dataset

amino acid	PS (rank)	AMAP (%)	non-AMAP (%)	difference (rank)	*p*-value
F-Phe	444(1)	10.05	4.81	5.24(1)	0.001
A-Ala	419(2)	11.23	7.45	3.78(2)	0.003
V-Val	403(3)	8.74	5.71	3.02(3)	0.030
H-His	382(4)	3.45	2.21	1.24(7)	0.051
W-Trp	372(5)	2.66	1.87	0.80(8)	0.126
Y-Tyr	371(6)	3.57	2.13	1.44(6)	0.101
M-Met	364(7)	1.19	1.26	−0.08(10)	0.846
P-Pro	364(8)	4.86	4.71	0.15(9)	0.847
K-Lys	359(9)	11.82	10.18	1.65(5)	0.203
D-Asp	355(10)	2.16	2.23	−0.08(11)	0.881
R-Arg	348(11)	5.90	6.06	−0.16(12)	0.884
C-Cys	341(12)	3.62	6.11	−2.49(18)	0.007
I-Ile	335(13)	4.64	6.63	−1.99(15)	0.006
L-Leu	333(14)	11.73	9.92	1.81(4)	0.178
T-Thr	320(15)	1.75	4.19	−2.44(17)	0.000
E-Glu	318(16)	0.84	2.20	−1.36(14)	0.000
N-Asn	318(17)	1.54	3.53	−1.99(16)	0.000
G-Gly	312(18)	7.10	10.62	−3.52(19)	0.009
Q-Gln	308(19)	1.80	2.23	−0.44(13)	0.343
S-Ser	275(20)	1.34	5.94	−4.60(20)	0.000

involved in peptide structural stability by restricting their conformation through hydrogen bond formation, affecting the antiplasmodial activity.[71] In addition, it has been reported that non-polar aliphatic amino acids (Ala, Ile, Leu, and Val) are responsible for the thermal stability of proteins and peptides.[72] These non-polar amino acid compositions indicate that AMAPs tend to be highly thermostable and conformationally stable in general. Thus, certain features such as high activity and thermostability coupled with peptide shelf life and stability in the host have made AMAPs an attractive endeavor for peptide drug discovery applications against malaria.

3.6.2. AMAPs Are Composed of Hydrophobic Amino Acids. The CIDH920102 property was found at rank five, which is described as normalized hydrophobicity scales for beta-proteins.[73] Thus, it could be implied that side-chain hydrophobicity might be important to AMAPs. Table 7 indicates the informative amino acids of the top five AMAPs (i.e., Phe, Ala, Val, His, and Trp) ranked by the propensity scores. In the meanwhile, the residues having the highest scores for the CIDH920102 property were Trp, Phe, Tyr, Ile, and Met.[73] The hydrophobicity of the nonpolar side of amphipathic helical peptides is correlated with peptide helicity and their ability to self-associate in aqueous environments, which is involved in peptide stability and membrane−peptide interaction.[72,74] An abundance of hydrophobic amino acids is significant in AMPs both of non-polar aliphatic and aromatic side chains.[75,76] Since a significantly high hydrophobic activity in AMPs can cause a strong hemolytic activity on erythrocytes,[74] an optimum hydrophobicity value is necessary to obtain both maximum specificity and activity of AMAPs.

3.6.3. AMAPs Are Composed of Amino Acids Interacting with Cell Membranes. To group amino acids based on peptide solubility, we analyzed the protein/peptide solubility based on the Amino Acid Index Database.[77] Relevant to the solvent accessibility property, the PCPs common to both the hydrophilic and hydrophobic residues, the BIOV880101 and BIOV880102 properties, are described as "Information value for accessibility; average fraction 35%" and "Information value for accessibility; average fraction 23%",[78] respectively. The

Table 7. Five Important Physicochemical Properties (PCPs) Derived from iAMAP-SCM[a]

amino acid	PS (rank)	PCP1 (rank)	PCP2 (rank)	PCP3 (rank)	PCP4 (rank)	PCP5 (rank)
F-Phe	444(1)	444(1)	23.0(2)	189(1)	2.9(1)	148(1)
A-Ala	419(2)	419(2)	9.8(11)	16(10)	1.3(8)	44(8)
V-Val	403(3)	403(3)	15.3(6)	123(7)	1.4(7)	117(7)
H-His	382(4)	382(4)	11.9(8)	50(9)	1.9(4)	47(4)
W-Trp	372(5)	372(5)	24.2(1)	145(4)	2.1(3)	163(3)
Y-Tyr	371(6)	371(6)	17.2(4)	53(8)	0.8(10)	22(10)
M-Met	364(7)	364(7)	11.9(9)	124(6)	2.8(2)	121(2)
P-Pro	364(8)	364(8)	15.0(7)	−20(12)	0.0(20)	−36(20)
K-Lys	359(9)	359(9)	10.5(10)	−141(20)	1.0(9)	−188(9)
D-Asp	355(10)	355(10)	4.9(14)	−78(18)	0.5(18)	−91(18)
R-Arg	348(11)	348(11)	7.3(12)	−70(14)	0.8(11)	−68(11)
C-Cys	341(12)	341(12)	3.0(17)	168(2)	0.7(12)	90(12)
I-Ile	335(13)	335(13)	17.2(3)	151(3)	1.6(5)	100(5)
L-Leu	333(14)	333(14)	17.0(5)	145(5)	1.4(6)	108(6)
T-Thr	320(15)	320(15)	6.9(13)	−38(13)	0.6(14)	−54(14)
E-Glu	318(16)	318(16)	4.4(15)	−106(19)	0.7(13)	−139(13)
N-Asn	318(17)	318(17)	3.6(16)	−74(17)	0.6(15)	−72(15)
G-Gly	312(18)	312(18)	0.0(20)	−13(11)	0.5(16)	−8(16)
Q-Gln	308(19)	308(19)	2.4(19)	−73(16)	0.2(19)	−117(19)
S-Ser	275(20)	275(20)	2.6(18)	−70(15)	0.5(17)	−60(17)
R	1.000	0.666	0.629	0.621	0.615	0.564

[a]PCP1 = TAKK010101 (side-chain contribution to protein stability (kJ/mol) (Takano−Yutani, 2001)); PCP2 = BIOV880101 (information value for accessibility; average fraction 35% (Biou et al., 1988)); PCP3 = RICJ880111 (relative preference value at C4 (Richardson−Richardson, 1988)); PCP4 = BIOV880102 (information value for accessibility; average fraction 23% (Biou et al., 1988)); PCP5 = CIDH92010 (normalized hydrophobicity scales for beta-proteins (Cid et al., 1992)).

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

high correlation between the 20 amino acid propensities and these two PCPs highlight that high membrane propensity is a key component of amino acids for AMAPs. Notably, the top seven amino acids with the highest propensity scores in the BIOV880101 and BIOV880102 properties (Phe, Cys, Ile, Trp, Leu, Met, and Val) are all non-polar side-chain amino acids.[78] These hydrophobic residues in the amphipathic AMPs affect membrane permeabilization via pore formation that is more sensitive to phospholipid bilayers with a negative charge.[79,80] This high positive correlation strongly indicates that AMAPs favor non-polar aliphatic and aromatic amino acids, which play an important role in peptide−membrane interactions.[81,82]

3.6.4. AMAPs Tend to be in the Amphipathic Antimicrobial Peptide Class. The property of RICJ880111, described as "relative preference value at C4",[83] was found at rank 3. As a result, the high positive correlation of the RICJ880111 property and the top eight informative amino acids (i.e., Phe, Met, Trp, Ile, Leu, Val, Ala, and His) indicate that AMAPs favor hydrophobic amino acids and tend to be in the amphipathic AMP class. The interface residue of globular protein structures, characterized by a half-in/half-out phenomenon, is termed N-cap or C-cap. This is based on the fact that the hydrophobic residues are highly conserved on α-carbon positions 4 of C-cap (C4).[83] The amphiphilic property enhances cellular membrane permeability through ion channel formation and thus disrupts the lipid bilayer.[84,85] In general, peptides acting on membranes exhibit their amphipathic activity in secondary structure form once they are in a membrane environment.[85] Owing to the potent clinical relevance of membrane active peptides, the correlation of AMP structure−function relationship has been extensively investigated.[85−88] The formation of amphipathic conformations can be observed when the cationic amphipathic peptides are located in membrane environments.[89−91] Notably, other conformations can also result in an amphipathic separation of hydrophobic and polar residues, efficient membrane interactions, and antimicrobial activities.[90,91] The amphipathic separation of hydrophobic and polar residues illustrates the specific interactions of AMPs with the phospholipid membrane interfaces.[89]

4. CONCLUSIONS

In this study, we have developed iAMAP-SCM, an interpretable and efficient ML-based predictor for the high-throughput identification and characterization of peptides with antimalarial activity using only sequence information. To the best of our knowledge, iAMAP-SCM is the first ML-based approach designed for AMAP identification and characterization. To be specific, this new predictor was developed based on the interpretable SCM method and estimated propensities of 20 amino acids and 400 dipeptides to be AMAPs. Experimental results based on the training and independent test datasets revealed that iAMAP-SCM achieved a competitive performance in AMAP prediction and outperformed several ML-based predictors in terms of simplicity and interpretability. Additionally, the iAMAP-SCM-derived analysis results demonstrated four important characteristics of AMAPs, which can be encapsulated as follows: (i) AMAPs have high conformational and thermal stability; (ii) AMAPs are composed of hydrophobic amino acids; (iii) AMAPs are composed of amino acids interacting with cell membranes; and (iv) AMAPs tend to be in the amphipathic antimicrobial peptide class. Finally, a user-friendly online computational platform of iAMAP-SCM is publicly available at http://pmlabstack.pythonanywhere.com/iAMAP-SCM.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c04465.

Performance evaluations of iAMAP-SCM and several ML-based classifiers; training and independent test datasets; peptide length distribution of AMAPs and non-AMAPs; hyperparameter search details; cross-validation results; independent test results; detailed information of the external dataset; detailed prediction results; and 20 top-ranked informative PCPs (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Pramote Chumnanpuen − *Department of Zoology, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand; Omics Center for Agriculture, Bioresources, Food, and Health, Kasetsart University (OmiKU), Bangkok 10900, Thailand*; Email: pramote.c@ku.th

Watshara Shoombuatong − *Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand*; orcid.org/0000-0002-3394-8709; Email: watshara.sho@mahidol.ac.th

Authors

Phasit Charoenkwan − *Modern Management and Information Technology, College of Arts, Media and Technology, Chiang Mai University, Chiang Mai 50200, Thailand*

Nalini Schaduangrat − *Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand*

Pietro Lio − *Department of Computer Science and Technology, University of Cambridge, Cambridgeshire CB3 0FD, U.K.*

Mohammad Ali Moni − *Artificial Intelligence & Digital Health, School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland, St Lucia QLD 4072, Australia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c04465

Notes

The authors declare no competing financial interest.

All the data used in this study are available at http://pmlabstack.pythonanywhere.com/iAMAP-SCM.

■ ACKNOWLEDGMENTS

This work was fully supported by College of Arts, Media and Technology, Chiang Mai University, and partially supported by Chiang Mai University and Mahidol University. In addition, computational resources were supported by Information Technology Service Center (ITSC) of Chiang Mai University.

■ REFERENCES

(1) Ghosh, A. K.; Ribolla, P. E.; Jacobs-Lorena, M. Targeting Plasmodium ligands on mosquito salivary glands and midgut with a

phage display peptide library. *Proc. Natl. Acad. Sci.* **2001**, *98*, 13278−13281.
(2) Gao, B.; Xu, J.; del Carmen Rodriguez, M.; Lanz-Mendoza, H.; Hernández-Rivas, R.; Du, W.; Zhu, S. Characterization of two linear cationic antimalarial peptides in the scorpion Mesobuthus eupeus. *Biochimie* **2010**, *92*, 350−359.
(3) Bell, A. Antimalarial peptides: the long and the short of it. *Curr. Pharm. Des.* **2011**, *17*, 2719−2731.
(4) Hancock, R. E. W.; Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* **2006**, *24*, 1551−1557.
(5) Mor, A. Multifunctional host defense peptides: antiparasitic activities. *FEBS J.* **2009**, *276*, 6474−6482.
(6) Splith, K.; Neundorf, I. Antimicrobial peptides with cell-penetrating peptide properties and vice versa. *Eur. Biophys. J.* **2011**, *40*, 387−397.
(7) Vizioli, J.; Bulet, P.; Hoffmann, J. A.; Kafatos, F. C.; Müller, H.-M.; Dimopoulos, G. Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector Anopheles gambiae. *Proc. Natl. Acad. Sci.* **2001**, *98*, 12630−12635.
(8) Shahabuddin, M.; Fields, I.; Bulet, P.; Hoffmann, J. A.; Miller, L. H. Plasmodium gallinaceum: differential killing of some mosquito stages of the parasite by insect defensin. *Exp. Parasitol.* **1998**, *89*, 103−112.
(9) Boman, H. G.; Wade, D.; Boman, I. A.; Wåhlin, B.; Merrifield, R. B. Antibacterial and antimalarial properties of peptides that are cecropin-melittin hybrids. *FEBS Lett.* **1989**, *259*, 103−106.
(10) Jaynes, J. M.; Burton, C. A.; Barr, S. B.; Jeffers, G. W.; Julian, G. R.; White, K. L.; Enright, F. M.; Klei, T. R.; Laine, R. A. In vitro cytocidal effect of novel lytic peptides on Plasmodium falciparum and Trypanosoma cruzi 1. *FASEB J.* **1988**, *2*, 2878−2883.
(11) Gwadz, R. W.; Kaslow, D.; Lee, J.-Y.; Maloy, W. L.; Zasloff, M.; Miller, L. H. Effects of magainins and cecropins on the sporogonic development of malaria parasites in mosquitoes. *Infect. Immun.* **1989**, *57*, 2628−2633.
(12) Moll, G. N.; van den Eertwegh, V.; Tournois, H.; Roelofsen, B.; den Kamp, J. A. O.; van Deenen, L. L. Growth inhibition of Plasmodium falciparum in in vitro cultures by selective action of tryptophan-N-formylated gramicidin incorporated in lipid vesicles. *Biochim. Biophys. Acta, Biomembr.* **1991**, *1062*, 206−210.
(13) Otten-Kuipers, M. A.; Roelofsen, B.; Op den Kamp, J. A. Stage-dependent effects of analogs of gramicidin A on the growth ofPlasmodium falciparum in vitro. *Parasitol. Res.* **1995**, *81*, 26−31.
(14) Gumila, C.; Ancelin, M.-L.; Delort, A.-M.; Jeminet, G.; Vial, H. J. Characterization of the potent in vitro and in vivo antimalarial activities of ionophore compounds. *Antimicrob. Agents Chemother.* **1997**, *41*, 523−529.
(15) McColm, A. A.; McHardy, N. Evaluation of a range of antimicrobial agents against the parasitic protozoa, Plasmodium falciparum, Babesia rodhaini and Theileria parva in vitro. *Ann. Trop. Med. Parasitol.* **1984**, *78*, 345−354.
(16) Bell, A.; Roberts, H. C.; Chappell, L. H. The antiparasite effects of cyclosporin A: possible drug targets and clinical applications. *Gen. Pharmacol.* **1996**, *27*, 963−971.
(17) Aminake, M. N.; Schoof, S.; Sologub, L.; Leubner, M.; Kirschner, M.; Arndt, H.-D.; Pradel, G. Thiostrepton and derivatives exhibit antimalarial and gametocytocidal activity by dually targeting parasite proteasome and apicoplast. *Antimicrob. Agents Chemother.* **2011**, *55*, 1338−1348.
(18) Clough, B.; Rangachari, K.; Strath, M.; Preiser, P. R.; Wilson, R. I. Antibiotic inhibitors of organellar protein synthesis in Plasmodium falciparum. *Protist* **1999**, *150*, 189−195.
(19) Sullivan, M.; Li, J.; Kumar, S.; Rogers, M. J.; McCutchan, T. F. Effects of interruption of apicoplast function on malaria infection, development, and transmission. *Mol. Biochem. Parasitol.* **2000**, *109*, 17−23.
(20) Boman, H. Antibacterial peptides: basic facts and emerging concepts. *J. Int. Med.* **2003**, *254*, 197−215.
(21) Nagaraj, G.; Uma, M. V.; Shivayogi, M. S.; Balaram, H. Antimalarial activities of peptide antibiotics isolated from fungi. *Antimicrob. Agents Chemother.* **2001**, *45*, 145−149.
(22) Aronimo, B. S.; Okoro, U. C.; Ali, R.; Ibeji, C. U.; Ezugwu, J. A.; Ugwu, D. I. Synthesis, molecular docking and antimalarial activity of phenylalanine-glycine dipeptide bearing sulphonamide moiety. *J. Mol. Struct.* **2021**, *1246*, No. 131201.
(23) Ekoh, O. C.; Okoro, U.; Ugwu, D.; Ali, R.; Okafor, S.; Ugwuja, D.; Attah, S. Novel Dipeptides Bearing Sulfonamide as Antimalarial and Antitrypanosomal Agents: Synthesis and Molecular Docking. *Med. Chem.* **2022**, *18*, 394−405.
(24) Ezugwu, J. A.; Okoro, U. C.; Ezeokonkwo, M. A.; Hariprasad, K. S.; Rudrapal, M.; Ugwu, D. I.; Gogoi, N.; Chetia, D.; Celik, I.; Ekoh, O. C. Design, Synthesis, Molecular Docking, Molecular Dynamics and In Vivo Antimalarial Activity of New Dipeptide-Sulfonamides. *ChemistrySelect* **2022**, *7*, No. e202103908.
(25) Mishra, M.; Singh, V.; Tellis, M. B.; Joshi, R. S.; Pandey, K. C.; Singh, S. Cyclic peptide engineered from phytocystatin inhibitory hairpin loop as an effective modulator of falcipains and potent antimalarial. *J. Biomol. Struct. Dyn.* **2022**, *40*, 3642−3654.
(26) Musyoka, T. M.; Njuguna, J. N.; Tastan Bishop, Ö. Comparing sequence and structure of falcipains and human homologs at prodomain and catalytic active site for malarial peptide based inhibitor design. *Malar. J.* **2019**, *18*, 1−21.
(27) Lhouvum, K.; Bhuyar, K. S.; Trivedi, V. Molecular modeling and correlation of PFI1625c-peptide models of bioactive peptides with antimalarial properties. *Med. Chem. Res.* **2015**, *24*, 1527−1533.
(28) Mehta, D.; Anand, P.; Kumar, V.; Joshi, A.; Mathur, D.; Singh, S.; Tuknait, A.; Chaudhary, K.; Gautam, S. K.; Gautam, A.; Varshney, G. C.; Raghava, G. P. S.ParaPep: a web resource for experimentally validated antiparasitic peptide sequences and their structures. *Database***2014**, *2014*, DOI: 10.1093/database/bau051.
(29) Silva, A. F.; Torres, M. D. T.; Silva, L.; Alves, F. L.; Pinheiro, A. A. D. S.; Miranda, A.; Capurro, M. L.; Oliveira, V. X. New linear antiplasmodial peptides related to angiotensin II. *Malar. J.* **2015**, *14*, 1−10.
(30) Mahindra, A.; Gangwal, R. P.; Bansal, S.; Goldfarb, N. E.; Dunn, B. M.; Sangamwar, A. T.; Jain, R. Antiplasmodial activity of short peptide-based compounds. *RSC Adv.* **2015**, *5*, 22674−22684.
(31) Choi, S.-J.; Parent, R.; Guillaume, C.; Deregnaucourt, C.; Delarbre, C.; Ojcius, D. M.; Montagne, J.-J.; Célérier, M.-L.; Phelipot, A.; Amiche, M.; Molgo, J.; Camadro, J. M.; Guette, C. Isolation and characterization of Psalmopeotoxin I and II: two novel antimalarial peptides from the venom of the tarantula Psalmopoeus cambridgei. *FEBS Lett.* **2004**, *572*, 109−117.
(32) Kokoza, V.; Ahmed, A.; Woon Shin, S.; Okafor, N.; Zou, Z.; Raikhel, A. S. Blocking of Plasmodium transmission by cooperative action of Cecropin A and Defensin A in transgenic Aedes aegypti mosquitoes. *Proc. Natl. Acad. Sci.* **2010**, *107*, 8111−8116.
(33) Krugliak, M.; Feder, R.; Zolotarev, V. Y.; Gaidukov, L.; Dagan, A.; Ginsburg, H.; Mor, A. Antimalarial activities of dermaseptin S4 derivatives. *Antimicrob. Agents Chemother.* **2000**, *44*, 2442−2451.
(34) Xiao, X.; Shao, Y.-T.; Cheng, X.; Stamatovic, B. iAMP-CA2L: a new CNN-BiLSTM-SVM classifier based on cellular automata image for identifying antimicrobial peptides and their functional types. *Brief. Bioinf.* **2021**, *22*, No. bbab209.
(35) Agrawal, P.; Bhagat, D.; Mahalwal, M.; Sharma, N.; Raghava, G. P. S. AntiCP 2.0: an updated model for predicting anticancer peptides. *Brief. Bioinf.* **2021**, *22*, No. bbaa153.
(36) Huang, H.-L.; Charoenkwan, P.; Kao, T.-F.; Lee, H.-C.; Chang, F.-L.; Huang, W.-L.; Ho, S.-J.; Shu, L.-S.; Chen, W.-L.; Ho, S.-Y. Prediction and analysis of protein solubility using a novel scoring card method with dipeptide composition. *BMC Bioinf.* **2012**, *13*, 1−14.
(37) Vasylenko, T.; Liou, Y.-F.; Chiou, P.-C.; Chu, H.-W.; Lai, Y.-S.; Chou, Y.-L.; Huang, H.-L.; Ho, S.-Y. SCMBYK: prediction and characterization of bacterial tyrosine-kinases based on propensity scores of dipeptides. *BMC Bioinf.* **2016**, *17*, 203−217.

(38) Charoenkwan, P.; Shoombuatong, W.; Lee, H.-C.; Chaijaruwanich, J.; Huang, H.-L.; Ho, S.-Y. SCMCRYS: predicting protein crystallization using an ensemble scoring card method with estimating propensity scores of P-collocated amino acid pairs. *PLoS One* **2013**, *8*, No. e72368.
(39) Liou, Y.-F.; Charoenkwan, P.; Srinivasulu, Y. S.; Vasylenko, T.; Lai, S.-C.; Lee, H.-C.; Chen, Y.-H.; Huang, H.-L.; Ho, S.-Y. SCMHBP: prediction and analysis of heme binding proteins using propensity scores of dipeptides. *BMC Bioinf.* **2014**, *15*, 1−14.
(40) Vasylenko, T.; Liou, Y.-F.; Chen, H.-A.; Charoenkwan, P.; Huang, H.-L.; Ho, S.-Y. SCMPSP: Prediction and characterization of photosynthetic proteins based on a scoring card method. *BMC Bioinf.* **2015**, 1−16.
(41) Charoenkwan, P.; Chiangjong, W.; Nantasenamat, C.; Moni, M. A.; Lio', P.; Manavalan, B.; Shoombuatong, W. SCMTHP: A New Approach for Identifying and Characterizing of Tumor-Homing Peptides Using Estimated Propensity Scores of Amino Acids. *Pharmaceutics* **2022**, *14*, 122.
(42) Charoenkwan, P.; Chiangjong, W.; Lee, V. S.; Nantasenamat, C.; Hasan, M. M.; Shoombuatong, W. Improved prediction and characterization of anticancer activities of peptides using a novel flexible scoring card method. *Sci. Rep.* **2021**, *11*, 1−13.
(43) Charoenkwan, P.; Chotpatiwetchkul, W.; Lee, V. S.; Nantasenamat, C.; Shoombuatong, W. A novel sequence-based predictor for identifying and characterizing thermophilic proteins using estimated propensity scores of dipeptides. *Sci. Rep.* **2021**, *11*, 1−15.
(44) Kawashima, S.; Pokarowski, P.; Pokarowska, M.; Kolinski, A.; Katayama, T.; Kanehisa, M. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res.* **2007**, *36*, D202−D205.
(45) Chen, Z.; Zhao, P.; Li, F.; Leier, A.; Marquez-Lago, T. T.; Wang, Y.; Webb, G. I.; Smith, A. I.; Daly, R. J.; Chou, K.-C.; Song, J. iFeature: a python package and web server for features extraction and selection from protein and peptide sequences. *Bioinformatics* **2018**, *34*, 2499−2502.
(46) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(47) Charoenkwan, P.; Chiangjong, W.; Nantasenamat, C.; Hasan, M. M.; Manavalan, B.; Shoombuatong, W. StackIL6: a stacking ensemble model for improving the prediction of IL-6 inducing peptides. *Brief. Bioinf.* **2021**, *22*, No. bbab172.
(48) Charoenkwan, P.; Nantasenamat, C.; Hasan, M. M.; Manavalan, B.; Shoombuatong, W. BERT4Bitter: a bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides. *Bioinformatics* **2021**, *37*, 2556−2562.
(49) Azadpour, M.; McKay, C. M.; Smith, R. L. Estimating confidence intervals for information transfer analysis of confusion matrices. *J. Acoust. Soc. Am.* **2014**, *135*, EL140−EL146.
(50) Lv, H.; Dao, F.-Y.; Guan, Z.-X.; Yang, H.; Li, Y.-W.; Lin, H. Deep-Kcr: accurate detection of lysine crotonylation sites using deep learning method. *Brief. Bioinf.* **2021**, *22*, No. bbaa255.
(51) Lv, H.; Dao, F.-Y.; Zulfiqar, H.; Su, W.; Ding, H.; Liu, L.; Lin, H. A sequence-based deep learning approach to predict CTCF-mediated chromatin loop. *Brief. Bioinf.* **2021**, No. bbab031.
(52) Wang, D.; Zhang, Z.; Jiang, Y.; Mao, Z.; Wang, D.; Lin, H.; Xu, D. DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism. *Nucleic Acids Res.* **2021**, *49*, e46−e46.
(53) Dao, F.-Y.; Lv, H.; Zulfiqar, H.; Yang, H.; Su, W.; Gao, H.; Ding, H.; Lin, H. A computational platform to identify origins of replication sites in eukaryotes. *Brief. Bioinf.* **2021**, *22*, 1940−1950.
(54) Dao, F.-Y.; Lv, H.; Zhang, D.; Zhang, Z.-M.; Liu, L.; Lin, H. DeepYY1: a deep learning approach to identify YY1-mediated chromatin loops. *Brief. Bioinf.* **2021**, *22*, No. bbaa356.
(55) Charoenkwan, P.; Kanthawong, S.; Nantasenamat, C.; Hasan, M. M.; Shoombuatong, W. iAMY-SCM: Improved prediction and analysis of amyloid proteins using a scoring card method with propensity scores of dipeptides. *Genomics* **2020**, 2813.
(56) Charoenkwan, P.; Kanthawong, S.; Nantasenamat, C.; Hasan, M. M.; Shoombuatong, W. iDPPIV-SCM: A sequence-based predictor for identifying and analyzing dipeptidyl peptidase IV (DPP-IV) inhibitory peptides using a scoring card method. *J. Proteome Res.* **2020**, *19*, 4125−4136.
(57) Chaianantakul, N.; Sungkapong, T.; Supatip, J.; Kingsang, P.; Kamlaithong, S.; Suwanakitti, N. Antimalarial effect of cell penetrating peptides derived from the junctional region of Plasmodium falciparum dihydrofolate reductase-thymidylate synthase. *Peptides* **2020**, *131*, No. 170372.
(58) Bianchin, A.; Bell, A.; Chubb, A. J.; Doolan, N.; Leneghan, D.; Stavropoulos, I.; Shields, D. C.; Mooney, C. Design and evaluation of antimalarial peptides derived from prediction of short linear motifs in proteins related to erythrocyte invasion. *PLoS One* **2015**, *10*, No. e0127383.
(59) Sinha, S.; Singh, A.; Medhi, B.; Sehgal, R. Systematic review: insight into antimalarial peptide. *Int. J. Pept. Res. Ther.* **2016**, *22*, 325−340.
(60) Leussa, A. N.-N.; Rautenbach, M. Antiplasmodial Cyclodecapeptides from Tyrothricin Share a Target with Chloroquine. *Antibiotics* **2022**, *11*, 801.
(61) Thongtan, J.; Saenboonrueng, J.; Rachtawee, P.; Isaka, M. An antimalarial tetrapeptide from the entomopathogenic fungus Hirsutella sp. BCC 1528. *J. Nat. Prod.* **2006**, *69*, 713−714.
(62) Ferreira, L. H. R.; Silva, A. F.; Torres, M. D. T.; Pedron, C. N.; Capurro, M. L.; Alves, F. L.; Miranda, A.; Oliveira, V. X. Effects of Amino Acid deletion on the antiplasmodial activity of angiotensin II. *Int. J. Pept. Res. Ther.* **2014**, *20*, 553−564.
(63) Liu, Z.; Brady, A.; Young, A.; Rasimick, B.; Chen, K.; Zhou, C.; Kallenbach, N. R. Length effects in antimicrobial peptides of the (RW) n series. *Antimicrob. Agents Chemother.* **2007**, *51*, 597−603.
(64) Tzakos, A. G.; Bonvin, A. M.; Troganis, A.; Cordopatis, P.; Amzel, M. L.; Gerothanassis, I. P.; Van Nuland, N. A. On the molecular basis of the recognition of angiotensin II (AII) NMR structure of AII in solution compared with the X-ray structure of AII bound to the mAb Fab131. *Eur. J. Biochem.* **2003**, *270*, 849−860.
(65) de Freitas, D. M.; Junior, V. X. O.Antiplasmodial Activity of Angiotensin II Analogs. *Proceedings of the 24 th American Peptide Symposium*; American Peptide Society, 2015.
(66) Dennison, S.; Harris, F.; Phoenix, D. A study on the importance of phenylalanine for aurein functionality. *Protein Pept. Lett.* **2009**, *16*, 1455−1458.
(67) Pérez-Picaso, L.; Velasco-Bejarano, B.; Aguilar-Guadarrama, A. B.; Argotte-Ramos, R.; Rios, M. Y. Antimalarial activity of ultra-short peptides. *Molecules* **2009**, *14*, 5103−5114.
(68) Mason, A. J.; Moussaoui, W.; Abdelrahman, T.; Boukhari, A.; Bertani, P.; Marquette, A.; Shooshtarizaheh, P.; Moulay, G.; Boehm, N.; Guerold, B.; Sawers, R. J. H.; Kichler, A.; Metz-Boutigue, M. H.; Candolfi, E.; Právost, G.; Bechinger, B. Structural determinants of antimicrobial and antiplasmodial activity and selectivity in histidine-rich amphipathic cationic peptides. *J. Biol. Chem.* **2009**, *284*, 119−133.
(69) Takano, K.; Yutani, K. A new scale for side-chain contribution to protein stability based on the empirical stability analysis of mutant proteins. *Protein Eng., Des. Sel.* **2001**, *14*, 525−528.
(70) Black, S. D.; Mould, D. R. Development of hydrophobicity parameters to analyze proteins which bear post-or cotranslational modifications. *Anal. Biochem.* **1991**, *193*, 72−82.
(71) Silva, A. F.; Bastos, E. L.; Torres, M. D. T.; Costa-da-Silva, A. L.; Ioshino, R. S.; Capurro, M. L.; Alves, F. L.; Miranda, A.; De Freitas Fischer Vieira, R.; Oliveira, V. X., Jr. Antiplasmodial activity study of angiotensin II via Ala scan analogs. *J. Pept. Sci.* **2014**, *20*, 640−648.
(72) Osorio, D.; Rondón-Villarreal, P.; Torres, R. Peptides: a package for data mining of antimicrobial peptides. *Small* **2015**, *12*, 44−444.

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095

(73) Cid, H.; Bunster, M.; Canales, M.; Gazitúa, F. Hydrophobicity and structural classes in proteins. *Protein Eng., Des. Sel.* **1992**, *5*, 373−375.
(74) Chen, Y.; Guarnieri, M. T.; Vasil, A. I.; Vasil, M. L.; Mant, C. T.; Hodges, R. S. Role of peptide hydrophobicity in the mechanism of action of α-helical antimicrobial peptides. *Antimicrob. Agents Chemother.* **2007**, *51*, 1398−1406.
(75) Pane, K.; Durante, L.; Crescenzi, O.; Cafaro, V.; Pizzo, E.; Varcamonti, M.; Zanfardino, A.; Izzo, V.; Di Donato, A.; Notomista, E. Antimicrobial potency of cationic antimicrobial peptides can be predicted from their amino acid composition: Application to the detection of "cryptic" antimicrobial peptides. *J. Theor. Biol.* **2017**, *419*, 254−265.
(76) Fjell, C. D.; Hiss, J. A.; Hancock, R. E. W.; Schneider, G. Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug. Discov.* **2012**, *11*, 37−51.
(77) Kawashima, S.; Kanehisa, M. AAindex: amino acid index database. *Nucleic Acids Res.* **2000**, *28*, 374−374.
(78) Biou, V.; Gibrat, J. F.; Levin, J. M.; Robson, B.; Garnier, J. Secondary structure prediction: combination of three different methods. *Protein Eng., Des. Sel.* **1988**, *2*, 185−191.
(79) Lehrer, R. I.; Barton, A.; Daher, K. A.; Harwig, S. S.; Ganz, T.; Selsted, M. E. Interaction of human defensins with Escherichia coli. Mechanism of bactericidal activity. *J. Clin. Invest.* **1989**, *84*, 553−561.
(80) Wimley, W. C.; Selsted, M. E.; White, S. H. Interactions between human defensins and lipid bilayers: evidence for formation of multimeric pores. *Protein Sci.* **1994**, *3*, 1362−1373.
(81) Piriou, F.; Lintner, K.; Fermandjian, S.; Fromageot, P.; Khosla, M. C.; Smeby, R. R.; Bumpus, F. M. Amino acid side chain conformation in angiotensin II and analogs: correlated results of circular dichroism and 1H nuclear magnetic resonance. *Proc. Natl. Acad. Sci.* **1980**, *77*, 82−86.
(82) Moll, G. N.; Konings, W. N.; Driessen, A. J. Bacteriocins: mechanism of membrane insertion and pore formation. *Antonie van Leeuwenhoek* **1999**, *76*, 185−198.
(83) Richardson, J. S.; Richardson, D. C. Amino acid preferences for specific locations at the ends of α helices. *Science* **1988**, *240*, 1648−1652.
(84) Kratochvil, H. T.; Newberry, R. W.; Mensa, B.; Mravic, M.; DeGrado, W. F. Spiers Memorial Lecture: Analysis and de novo design of membrane-interactive peptides. *Faraday Discuss.* **2021**, *232*, 9−48.
(85) Dathe, M.; Wieprecht, T.; Nikolenko, H.; Handel, L.; Maloy, W. L.; MacDonald, D. L.; Beyermann, M.; Bienert, M. Hydrophobicity, hydrophobic moment and angle subtended by charged residues modulate antibacterial and haemolytic activity of amphipathic helical peptides. *FEBS Lett.* **1997**, *403*, 208−212.
(86) Maloy, W. L.; Kari, U. P. Structure−activity studies on magainins and other host defense peptides. *Biopolymers* **1995**, *37*, 105−122.
(87) Zhong, L.; Putnam, R. J.; Johnson, J. R.; Rao, A. G. Design and synthesis of amphipathic antimicrobial peptides. *Int. J. Pept. Protein Res.* **1995**, *45*, 337−347.
(88) Sitaram, N.; Subbalakshmi, C.; Nagaraj, R. Structural and charge requirements for antimicrobial and hemolytic activity in the peptide PKLLETFLSKWIG, corresponding to the hydrophobic region of the antimicrobial protein bovine seminalplasmin. *Int. J. Pept. Protein Res.* **1995**, *46*, 166−173.
(89) Bechinger, B. Rationalizing the membrane interactions of cationic amphipathic antimicrobial peptides by their molecular shape. *Curr. Opin. Colloid Interface Sci.* **2009**, *14*, 349−355.
(90) Zasloff, M. Antimicrobial peptides of multicellular organisms. *Nature* **2002**, *415*, 389−395.
(91) Hwang, P. M.; Vogel, H. J. Structure-function relationships of antimicrobial peptides. *Biochem. Cell Biol.* **1998**, *76*, 235−246.

https://doi.org/10.1021/acsomega.2c04465
ACS Omega 2022, 7, 41082−41095