UNIVERSIDAD REGIONAL AMAZÓNICA IKIAM
FACULTAD DE CIENCIAS DE LA VIDA
CARRERA EN BIOTECNOLOGÍA

DESIGN OF POTENTIAL LEISHMANICIDAL PEPTIDES ASSISTED BY ARTIFICIAL INTELLIGENCE

Proyecto de investigación previo a la obtención del Título de:
Ingeniero en Biotecnología

AUTOR: ALBERTO ALEXANDER ROBLES LOAIZA
TUTOR: PHD. RAFAEL DE ALMEIDA
COTUTORA: PHD. PATRICIA SALERNO

Napo, Ecuador
2022

DECLARACIÓN DE DERECHO DE AUTOR

Yo, Alberto Alexander Robles Loaiza, con documento de identidad N° 0706422748, declaro que los resultados obtenidos en la investigación que presento en este documento final, previo a la obtención del título de Ingeniería en Biotecnología, son absolutamente inéditos, originales, auténticos y personales.

En virtud de lo cual, el contenido, criterios, opiniones, resultados, análisis, interpretaciones, conclusiones, recomendaciones y todos los demás aspectos vertidos en la presente investigación son de mi autoría y de mi absoluta responsabilidad.

Tena, 13 de julio de 2022

Alberto Alexander Robles Loaiza
0706422748

AUTORIZACION DE PUBLICACION EN EL REPOSITORIO INSTITUCIONAL

Yo, Alberto Alexander Robles Loaiza, con documento de identidad N° 0706422748, en calidad de autor/a y titular de los derechos morales y patrimoniales del trabajo de titulaciòn: Design of potential leishmanicidal peptides assisted by artificial intelligence de conformidad con el Art. 114 del CÒDIGO ÒRGANICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÒN, reconozco a favor de la Universidad Regional Amazónica Ikiam una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Así mismo autorizo a la Universidad Regional Amazònica Ikiam para que realice la publicación de este trabajo de titulaciòn en el Repositorio Institucional de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación superior.

Tena, 13 de julio de 2022

Alberto Alexander Robles Loaiza
0706422748

CERTIFICADO DE DIRECCIÓN DE TRABAJO DE INTEGRACIÓN CURRICULAR

Certifico que el Trabajo de Integración Curricular Titulado: Design of potential leishmanicidal peptides assisted by artificial intelligence, en la modalidad: artículo, fue realizado por: Alberto Alexander Robles Loaiza, bajo mi dirección.

El mismo ha sido revisado en su totalidad y analizado por la herramienta de verificación de similitud de contenido; por lo tanto, cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad Regional Amazónica Ikiam, para su entrega y defensa.

Tena, 13 de julio de 2022

Firma
José Rafael de Almeida
1757448954

Patricia Elena Salerno
1759267857

DEDICATORIA

Este trabajo es dedicado a mis dos tías y abuela que padecen de esquizofrenia.

Es el mejor gesto que puedo hacer para que el mundo sepa que también contribuyeron en esta sociedad. Ellas en gran medida me enseñaron lo que significa el valor de la vida y mediante estas pocas, pero valiosas palabras espero que todo aquel que lea mi trabajo, sepa que una persona con “discapacidad” en realidad es un ser humano con capacidades especiales.

Sin ellas, no supiera lo que es ser un loco apasionado en un mundo de cuerdos. Sin Jessica, Gina y Esperancita, probablemente este trabajo no tuviera razón de ser.

AGRADECIMIENTO

Mi más sincero agradecimiento al docente Rafael de Almeida por compartir toda su experticia y ser un apoyo invaluable e incondicional antes y durante el desarrollo y ejecución de este trabajo de titulación.

A Patricia Salerno y Fabien Plisson por contribuir con su experiencia para darle moldura a esta tesis de pregrado.

A todos los docentes, mis padres y amigos que fueron indispensables para alcanzar esta etapa de ser un profesional.

TABLA DE CONTENIDOS

INDICE DE TABLAS

INDICE DE FIGURAS

RESUMEN

Una de las enfermedades tropicales desatendidas que han puesto en jaque al mundo por tener terapias convencionales tóxicas, poca inversión para el descubrimiento de nuevos fármacos y resistencia emergente es la leishmaniasis. En base a este contexto, la presente investigación tuvo como objetivo emplear por primera vez algoritmos de inteligencia artificial para ayudar en el diseño *in silico* de posibles péptidos leishmanicidas. Para ello se utilizaron 10 técnicas de machine learning y una optimización de los modelos utilizando matrices de correlación de Kendall. Estas herramientas se entrenaron en un conjunto de datos curado que consiste en péptidos antileishmania (ALPs), antimicrobianos (AMPs) y no parasitarios (no APPs) probados experimentalmente y almacenados en la literatura: ADP3 y PredAPP. Luego, se diseñaron y testearon 5000 secuencias aleatorias, de las cuales 221 se destacaron como posibles agentes que podrían presentar acción contra *Leishmania*. Todas estas moléculas con actividad potencial son nuevas porque no se encuentran en las bases de datos de péptidos antimicrobianos y parasitarios más actualizadas. La exactitud de clasificación de los 5 modelos establecidos es: Random Forest (RF): 92%, Support Vector Machines con Polynomial Kernel: 90%, Stochastic Gradient Boosting (SGB): 89% para la clasificación de no APP - ALP y RF - 87%, SVM P - 85% para la predicción de clases AMP - ALP. Por lo tanto, con esta investigación se presenta un útil enfoque *in silico* basado en inteligencia artificial que permite el descubrimiento a gran escala de péptidos antileishmania.

PALABRAS CLAVES: Leishmaniasis, Inteligencia artificial, Machine learning, Péptidos antileishmania, Péptidos antimicrobianos, péptidos antiparasitarios.

ABSTRACT

One of the neglected tropical diseases that have put the world in check for having toxic conventional therapies, low investment for the discovery of new drugs and emerging resistance is leishmaniasis. Based on this context, the present investigation aimed to employ for the first-time artificial intelligence algorithms to aid in the *in silico* design of potential leishmanicidal peptides. For this purpose, 10 machine learning techniques and an optimization of the models using Kendall correlation matrices were used. Our models were trained on a curated dataset consisting of experimentally-tested leishmanicidal (ALPs), antimicrobial (AMPs) and non-parasitic peptides (non-APPs) and stored in literature: ADP3 and PredAPP. Then, 5000 random amino acid sequences were designed and tested, of which 221 stood out as potential antileishmanial agents. All of these molecules with potential activity are novel because they are not found in the most up-to-date antimicrobial and parasitic peptide databases. The accuracy prediction percentages of the 5 models established are: Random Forest (RF): 92%, Support Vector Machines with Polynomial Kernel: 90%, Stochastic Gradient Boosting (SGB): 89% for the classification of non-APP - ALP and RF - 87%, SVM P - 85% for the prediction of AMP - ALP classes. Therefore, a useful in silico approach based on artificial intelligence that enables large-scale discovery of antileishmanial peptides is presented.

KEYWORDS: antimicrobial peptides, antiparasitic peptides, antileishmanial peptides, intelligence artificial, leishmaniasis, machine learning.

INTRODUCTION

Leishmaniasis is characterized as an ancient disease of vulnerable regions. According to the World Health Organization (WHO), this neglected tropical challenge is caused by the *Leishmania* parasite [1] and transmitted by vectors of the genus *Phlebotomus* and *Lutzomyia* [2]. Malnutrition, displacement, poor housing conditions, weak immune systems and lack of economic resources are generally risk factors associated with this pathology [3, 4]. There are an estimated 0.9 to 1.6 million cases, 20,000 to 30,000 deaths and 350 million people at risk of infection each year, making it the second most deadly parasitic disease after malaria [5].

Although leishmaniasis has been recognized for more than a century, its therapeutic arsenal such as pentavalent antimonials, amphotericin B and miltefosine have not been able to fully treat the disease [6]. The number of side effects [7], emergence of resistant strains [8, 9], intracellular nature and low investment in research has made it difficult to mitigate its impact [6]. As consequence, the discovery and development of alternative chemotherapies are a priority.

Synthetic or natural peptides have been highlighted as attractive scaffolds of high interest [10]. These short molecules combine advantageous features for clinical translational, with several successful examples approved by FDA. More than 200 peptide-derived drugs are commercially available [11]. Motivated by this, many studies have demonstrated that *in vitro* and *in vivo* antileishmanial activity of peptides, offering new chemical structures for the design of the next-generation of antiparasitic agents [12]. However, the peptide development pipeline remains laborious, time-consuming, costly, and with many challenges.

In past few years, advances in computational area have revolutionized the high-throughput screening of peptide activity. Many a database-driven bioinformatics based on machine and deep learning approaches have been introduced to predict the therapeutic effects of peptides. They have positioned as valuable strategies to identify optimal candidates to the chemical synthesis and functional testing in wet-lab [13]. Most of these techniques have focused on anticancer [14], hemolytic [15], antimicrobial [16], and antiviral [17] activities. For instance, there are more than 8 algorithms developed to predict the hemolytic activity of these molecules [15]. In contrast, a small portion (only

one) of these *in silico* programs have explored the antiparasitic properties of peptides [18].

To date, as far as our knowledge, no computational algorithm has been developed to aid for discovering of peptides targeting *Leishmania* parasites. Therefore, the aim of this study is to employ for the first-time artificial intelligence algorithms for the *in silico* design of leishmanicidal peptides. This tool translates into a shortest and cheapest shortcut to discovering novel potential candidates to treat an epidemiological relevant, but neglected tropical diseases [19].

METHODS

Development of AMP – ALP and non-APP – ALP classification models

Preprocessing of AMPs, non-APPs and ALPs databases

Figure 1 illustrates the general workflow proposed in this investigation to screen and design antileishmanial peptides assisted by artificial intelligence. Three peptide databases containing only natural amino acids were employed: APD3 (3166 AMP sequences), PredAPP (1890 non-APPs) and the database recently described by Robles-Loaiza et. al. with 110 ALPs [12, 15]. Firstly, based on the amino acid sequence of each AMP, non-AMP and ALP peptides, 56 physico-chemical descriptors, amino acid frequencies (20 variables), dimers (400 variables) and trimers (8000 variables) were analyzed by ModlAMPs package [20] implemented in Python (version 3.9.7) and the Rcpi package [21] in R (version 4.1.2). Scripts are available at https://github.com/albert-robles1101/Design-of-ALPs-assisted-by-artificial-intelligence. Overall, 8476 variables were explored.

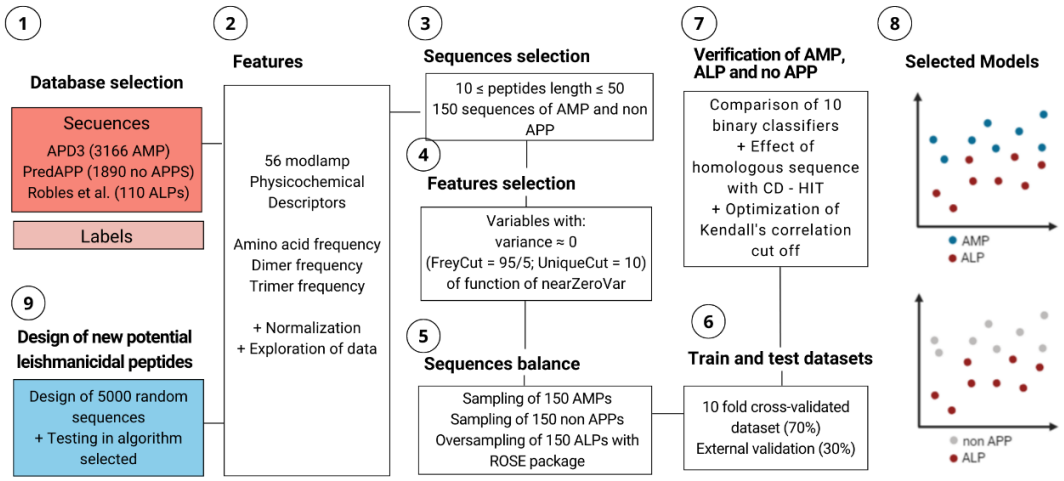

Figure 1. Overview of the classification models architecture.
Realizado por: Robles, Alberto, 2022

From left to right, (1) Selection of AMPs, non APPs and ALPs databases, (2) Determination, normalization and exploration of 8476 variables: 56 physical-chemical descriptors; 20 for the frequency of amino acids; 400 for the frequency of dimers and 8000 for the frequency of trimers, (3) Selection of 150 AMPs and non-APPs sequences that have between 10 and 50 lengths in their amino acid residues, (4) Elimination of non-significant variables (variance ≈ 0) for the elaboration of the models, (5) Balancing of leishmanicidal sequences using oversampling of the ROSE package to have balanced datasets: n AMPs = n ALPs and n non-APPs = n ALPs, (6) Elaboration of the training database and test, (7) Development of 10 binary classifiers (AMPs vs ALPs and nonAPPs vs ALPs) including analysis of homologous leishmanicidal peptide sequences and model optimization, (8) Selection of the models based on those with the best accuracy and avoiding overfitting and underfitting of the data and (9) Design of 5000 random sequences and test in the selected algorithms.

Only peptide sequences having 10 to 50 amino acid residues were selected. Subsequently, pseudo-random sampling was performed to obtain 150 antimicrobial peptides of APD3 and 150 non-antiparasitic peptides from the PredAPP database. This sampling was then balanced against the 110 antileishmanial peptide sequences in the ALP database. As a result, two data sets composed of: 1.- the merger of the ALP database with AMP – 150; 2.- the ALP dataset with non – APP – 150 were generated. In addition, all non-significant variables that had variance around to zero were removed based on the parameters freqCut = 95/5 and uniquecut = 10 of the nearZeroVar function in both databases.

An oversampling was performed using the ovun.sample function of the ROSE package [22] in R aiming to balance the databases with 150 observations for each group. In summary, the ALP – AMP 150 database contains 300 observations and 98 variables, while the ALP – non – APP 150 database is composed by 300 observations and 88 variables. Finally, a pseudo-random sampling was performed, and the data were split into training and test dataset: 70 % of the ALP - AMP 150 and ALP - non-APP 150 and 30 %, respectively.

Data exploration

An exploratory analysis of the ALP - AMP 150 and ALP - non-APP 150 databases was performed. The mean, standard deviation, 25th quartile, 50th, 75th quartile, skewness, kurtosis, univariate and multivariate normality were calculated for the 56 physicochemical properties using the MVN Package of R. In addition, Pearson's and Kendall's correlation matrices were determined and dimensionality reduction techniques were applied to understand how each of the variables to be studied would behave.

Development, validation and selection of the AMP – ALP and non-APP – ALP classification models

Ten models were chosen Logistic Regression, Naive Bayes, Bagged CART, Random Forest, Support Vector Machines with Linear Kernel, Support Vector Machines with Polynomial Kernel, Support Vector Machines with Radial Basis Function Kernel, Stochastic Gradient Boosting, Quadratic Discriminant Analysis with Step-by-step feature selection, linear discriminant analysis; and implemented in “train” function of the Caret package in R. These algorithms are the mostly employed for peptide activity prediction [23-25]. A K= 10-fold cross-validation procedure was run to analyze performance. Accuracy, KAPPA, Precision, Recall, F1, MCC, AUC – ROC were estimated. The definitions of these metrics are detailed in the supplementary material (**Supplemental definitions**). For the validation of the models, the test sets of the ALP – AMP 150 and ALP – non-APP 150 databases were used. Likewise, the aforementioned evaluation parameters were calculated, and the training and testing results were compared. Finally, the best AMP-ALP and non-APP-ALP classification models were selected trying to avoid overfitting or underfitting in the training and test data.

Effect of homologous sequences on classification models

The influence of the homologous sequences that may exist between the leishmanicidal peptides and the classification models were analyzed using the different CD-HIT homology cutoffs: 1, 0.90, 0.80, 0.70, 0.60, 0.50 [26-29] and the accuracy evaluation metric. This means that all the algorithms described in 2.1.3 were developed, where previously a balancing of the two datasets: AMP - ALP and nonAPP - ALP was performed after having performed the respective antileishmanial peptides identity cuts mentioned above. For the balancing and selection of training and test data, the process detailed in the third paragraph of 2.1.1 was repeated.

Optimization of the AMP – ALP and non-APP – ALP classification models based on Kendall's correlation cutoffs

To optimize the AMP – ALP and non-APP – ALP classification models, the Kendall correlation cutoffs were determined: 1, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70. Pearson's correlation cutoff was not used because in the exploratory data analysis no differences were visualized in the AMP, non-APP and ALP groups when applying linear statistical techniques. Accuracy was compared and the models with the best evaluation metrics were selected according to the correlation cutoff. An algorithm is considered "best" when its accuracy is high and very similar for both training and test data, which means that there will be no bias of the results due to overfitting or underfitting.

Design of new potential leishmanicidal peptides

5000 random sequences having a length of 10 to 30 amino acid residues were designed using the generate_sequences function of the Python modlAMP package. Posteriorly, 56 physicochemical properties, amino acid frequencies, dimers and trimers were calculated for each one of the peptides. The antileishmanial potential of these designed molecules was determined using the five selected algorithms. Based on the research of Plisson et. al. [30], established as probable when the average probability of the results of the 5 models coincided with the value of the ALP class (average probability > 0.5). Subsequently, a heatmap in R and structural predictions were performed using PEP2D [31] of the characterized peptides with leishmanicidal activity.

RESULTS AND DISCUSION

Data quality, behavior, and variable independence

The data's quality and quantity, the variable's independence and the type of algorithm are important factors in the predictive or classification capacity of artificial intelligence models [30, 32]. Motivated by this, in the present study, we employed 10 models useful for predicting antimicrobial, non-antiparasitic and antileishmanial activity of the peptides. Information about the non-antileishmanial activity of peptides is scarce in literature, hindering the design of *in silico* tools for this specific prediction. As has been implemented in other research for the development of *in silico* tools, the algorithms proposed in this study were developed under the assumption that peptides could be classified based on their physicochemical properties and frequencies of amino acid, dimer and trimer [33, 34]. To this end, the number of variables was reduced from 8476 to 98 dataset of AMP - ALP and from 8476 to 88 in the case of non-APP – ALP. This approach looks for useful and representative variables for algorithm implementation. Consequently, the minimum memory requirements are lower and the data processing is faster.

The physicochemical descriptors of AMPs, APPs and ALPs did not exhibit a multivariate normality. Of the 56 evaluated properties, 14, 3 and 3 showed normality for antimicrobial, non-antiparasitic and antileishmanial peptides, respectively. The means, standard deviations, median, minimum, maximum, 25th and 75th quartiles, skewness, and kurtosis are presented in detail in supplementary material (**Table S1**, **Table S2**, and **Table S3**). The physicochemical descriptors, as well as the Pearson and Kendall relationships of each of the variables, were key for the development and optimization of the classification models.

The dimensionality reduction techniques (**Figure 2 A** and **Figure 2 B**): DRR, KamadaKawai, Isomap, PCA_L1, MDA, nMDA, PCA, tSNE; do not manage to have a good quality measure of embedding of the data according to the R_{NX} score, therefore, the AMP, nonAPP and ALP classes cannot be distinguished in two groups (AMP - ALP or nonAPP - ALP). This convergence of the clusters can be observed, as an example using PCA, in **Figure 2 C** for the AMP - ALP and **Figure 2 D** for nonAPP - ALP groups. PCA is a tool that is based on Pearson correlations [35] and clearly as can be noted it is not particularly useful for these data. Based on this antecedent, the use of Kendall

correlations compared to Pearson correlations for model optimization is also justified. Likewise, it is shown that to classify AMP, non-APP and ALP datasets, non-linear statistical tools are needed, as is the case of the proposed algorithms. **Figure S1 – S7** of the supplementary material shows the Pearson and Kendall correlation matrices for the AMP, nonAPP and APP datasets, as well as the convergences of the groups to be classified by the different dimensionality reduction techniques.

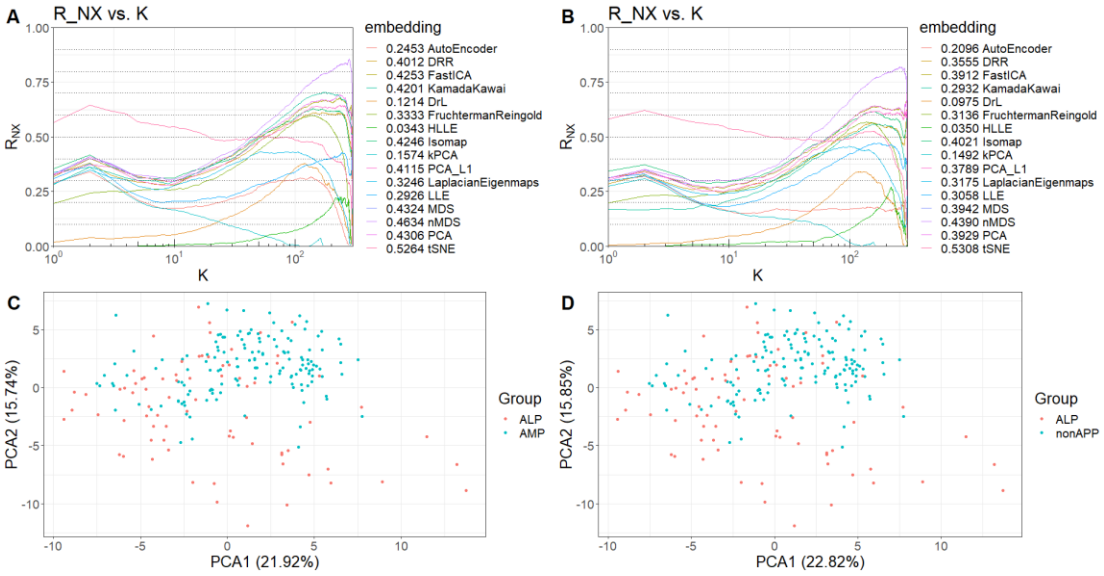

Figure 2. R_{NX} quality measure according to the size of the neighborhood of the K value of the dimensionality reduction methods for the classification models.

Realizado por: Robles, Alberto, 2022

A: R_{NX} versus K for the AMP-ALP database. **B:** R_{NX} vs K for non-APP – ALP database. **C:** Exemplification of the PCA dimensionality reduction technique for AMP – ALP database. **D:** PCA for non-APP - ALP database.

Model evaluation and improvement

Table 1 and **Table 2** evidenced that the non-APP – ALP classification models have better performance evaluation metrics than the AMP – ALP algorithms [18, 23, 36]. The accuracy for the best non-APP – ALP classification models are around 88 – 90% for the training data and 84 – 88% for the test set. For its part, the AMP - ALP training dataset had an accuracy of 83 to 88%, while for the test dataset it performed from 78 to 86%. The algorithms that stood out for their performance were: Random Forest (RF), Support Vector Machine with Polynomial Kernel (SVM P) and Stochastic Gradient Boosting

(SGB). SVM, RF and CART are algorithms that also in previously reported AMP classification models such as: AMPfun [23], ClassAMP [24] y iAMPred [25].

Table 1. Evaluation metrics of the 10 models developed to predict the AMP or ALP class.

Class.	Model (105/105)						Evaluation (45/45)					
	Acc.	Prec	Recall	F1	MCC	AUC - ROC	Acc.	Prec	Recall	F1	MCC	AUC - ROC
Logit	0.76	0.72	0.84	0.77	0.52	0.76	0.67	0.64	0.76	0.69	0.34	0.67
NB	0.79	0.76	0.84	0.80	0.48	0.74	0.77	0.72	0.87	0.79	0.54	0.77
B CART	0.83	0.82	0.86	0.84	0.67	0.83	0.78	0.72	0.91	0.80	0.58	0.78
RF	**0.88**	0.85	0.90	0.87	0.72	0.86	**0.83**	0.79	0.91	0.84	0.67	0.83
SVM L	0.79	0.77	0.82	0.79	0.57	0.79	0.66	0.62	0.78	0.69	0.32	0.66
SVM P	**0.85**	0.84	0.87	0.85	0.54	0.77	**0.86**	0.82	0.91	0.86	0.72	0.86
SVM RBFK	0.78	0.77	0.81	0.79	0.52	0.76	0.79	0.75	0.87	0.80	0.58	0.79
SGB	0.84	0.83	0.87	0.85	0.63	0.82	0.77	0.71	0.89	0.79	0.55	0.77
QDA	0.64	0.62	0.71	0.67	0.29	0.64	0.60	0.59	0.64	0.62	0.20	0.60
LDA	0.71	0.67	0.84	0.75	0.44	0.71	0.63	0.60	0.80	0.69	0.28	0.63

Realizado por: Robles, Alberto, 2022

Ninety-eight variables and a training and test data ratio of 0.7/0.3, respectively, were used. Accuracies underlined in bold highlight the algorithms with the best performance metrics for both the training and test dataset. RF and SVM P are potential tools for classification because they exhibit minimal overfitting and underfitting on the datasets. The acronyms of the developed models are: Logistic Regression (Logit), Naive Bayes (NB), Bagged CART (B CART), Random Forest (RF), Support Vector Machines with Linear Kernel (SVM L), Support Vector Machines with Polynomial Kernel (SVM P), Support Vector Machines with Radial Basis Function Kernel (SVM RBFK), Stochastic Gradient Boosting (SGB), Quadratic Discriminant Analysis with Step-by-step feature selection (QDA) and linear discriminant analysis (LDA).

Table 2. Accuracy, Precision, Recall, F1 – score, MCC and AUC - ROC of the 10 classification models used to predict non-APPs and ALPs.

Class.	Model (105/105)						Evaluation (45/45)					
	Acc.	Prec	Recall	F1	MCC	AUC - ROC	Acc.	Prec	Recall	F1	MCC	AUC - ROC
Logit	0.73	0.70	0.83	0.76	0.48	0.73	0.80	0.75	0.91	0.82	0.62	0.80
NB	0.78	0.80	0.75	0.77	0.56	0.78	0.79	0.74	0.89	0.81	0.59	0.79
B CART	0.88	0.85	0.92	0.86	0.76	0.88	0.83	0.81	0.87	0.84	0.67	0.83
RF	**0.89**	0.87	0.92	0.89	0.77	0.89	**0.88**	0.81	0.98	0.89	0.77	0.88
SVM L	0.76	0.73	0.82	0.77	0.52	0.76	0.79	0.74	0.89	0.81	0.59	0.79
SVM P	**0.90**	0.89	0.91	0.90	0.61	0.81	**0.88**	0.90	0.84	0.87	0.76	0.88
SVM RBFK	0.83	0.81	0.88	0.84	0.65	0.83	0.82	0.77	0.91	0.84	0.65	0.82
SGB	0.89	0.85	0.94	0.86	0.74	0.87	0.84	0.78	0.96	0.86	0.71	0.84
QDA	0.71	0.74	0.68	0.71	0.44	0.72	0.63	0.64	0.62	0.63	0.27	0.63
LDA	0.75	0.72	0.82	0.76	0.50	0.75	0.77	0.71	0.91	0.80	0.56	0.77

Realizado por: Robles, Alberto, 2022

Eighty-eight variables and a training and test data ratio of 0.7/0.3, respectively, had to be used during the implementation of the algorithms. The algorithms with best evaluation metrics for predicting are highlighted in bold. RF and SVM P are potentially viable algorithms because they do not overfit or underfit both training and test data. The acronyms of the models are: Logistic Regression (Logit), Naive Bayes (NB), Bagged CART (B CART), Random Forest (RF), Support Vector Machines with Linear Kernel (SVM L), Support Vector Machines with Polynomial Kernel (SVM P), Support Vector Machines with Radial Basis Function Kernel (SVM RBFK), Stochastic Gradient Boosting (SGB), Quadratic Discriminant Analysis with Step-by-step feature selection (QDA) and linear discriminant analysis (LDA).

On the models, different homology cuts off were made in the CD-HIT clusters. **Figure 3** shows the different accuracy values (ACC) of four developed classification models: BT, RF, SGB and SVM P, which are the ACC for the AMP - ALP classification models. There is an underfitting when there is a cut off 0.90 in the sequences. On the other hand, we observed overfitting with the remaining homology cuts: 1, 0.80, 0.70, 0.60, 0.50. Finally, **C** and **D**, which correspond to the ACC of the non-APP – ALP classification algorithms, alternate overfitting and underfitting are obtained for each CD – HIT redundancy elimination cut. By far, this is the first investigation that take account the selection of antileishmanial sequences according to the identity. However, CD-HIT is often used in

model architecture of the AMP activity prediction models [37, 38]. Temporins [39, 40], Bombins[41], Cruzioseptins [42], Phylloseptins [43-45], Dermaseptins [46, 47], Protamins [12], Magainins [12, 48] possibly influence the accuracy of models. In short words, removing redundant sequences was not particularly helpful in reducing bias in the predictions. In this context, we deploy the Kendall correlation bounds. The ACC of the additional models according to the percent identity are shown in supplementary material (**Table S4** and **Table S5)**.

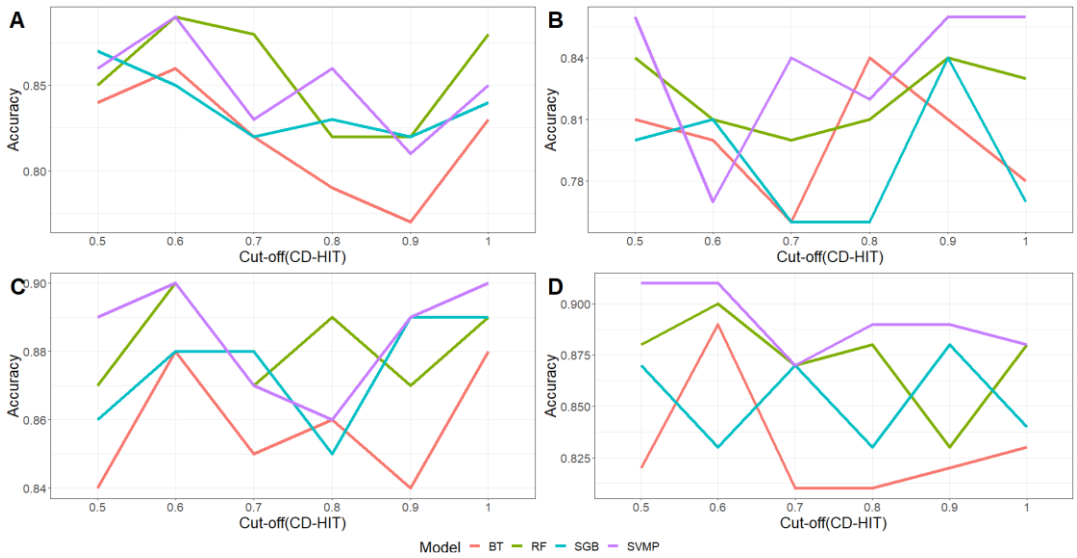

Figure 3. Representation of the accuracy values of classification models according to CD-HIT homology cut-off.
Realizado por: Robles, Alberto, 2022

A and **B** relate to AMP - ALP prediction of training and test data respectively. C and D correspond to nonAPP - ALP training and test data, respectively. The names of models are: Bagged CART (BT), Random Forest (RF), Stochastic Gradient Boosting and Support Vector Machines with Polynomial Kernel (SVM P).

Optimization was performed for the following Kendall correlation cutoff points: 0.75, 0.70, 080, 0.85, 0.90, 0.95, 1. Pearson correlation cutoff points were not used in the optimization models because, as mentioned above, the behavior of the variables in general is nonlinear. The correlation cuts allowed eliminating multicollinearity and ensuring that each variable had the best quality and representativeness in the models [30, 49].

Table 3 and **Table 4** show the accuracy of the AMP – ALP and non-APP – ALP models, respectively, with the different correlation cutoffs. For a Kendall correlation (corr) <- 0.75, the best accuracy results were obtained from the AMP - ALP classification models, where 86 of the 98 initial variables were used. The featured RF and SVM P models presented an accuracy value of 85 – 87% for the training data and 84% for the test set. On the other hand, the most optimal correlation cutoff for the non-APP – ALP classification models were 0.70 with the use of 70 of the 88 initial variables. The featured models RF, SVM P and SGB presented an accuracy between 89 – 92% for the training data and 88 – 89% for the test set. In general, this way of optimizing the models successfully eliminated the overfitting or underfitting that was present in the data. On the other hand, compared to other models in the literature, our algorithms presented better evaluation metrics. For example, AMPfun is a bioinformatics tool that classifies antiparasitic peptides with antimicrobial peptides and other types of activity. The accuracy for prediction of antiparasitic activity was 83.21% based on CART algorithm, while ours was 85 - 87% based on Support Vector Machine Polynomial and Random Forest, respectively [23]. Likewise, PredAPP classified parasitic and non-parasitic peptides with an accuracy of 88% [18]. We had a performance of 89, 90, 92% in Stochastic Gradient Boosting, Support Vector Machine Polynomial and Random Forest algorithms, respectively.

Table 3. Optimization of the AMP - ALP prediction models based on the Kendall correlation cuts-off.

Corr	Model (105/105)							Evaluation (45/45)						
	1	0.95	0.9	0.85	0.8	0.75	0.7	1	0.95	0.9	0.85	0.8	0.75	0.7
N	98	96	95	93	92	86	81	98	96	95	93	92	86	81
LR	0.76	0.73	0.69	0.74	0.71	0.77	0.75	0.67	0.69	0.69	0.69	0.67	0.68	0.74
NB	0.79	0.79	0.76	0.78	0.76	0.78	0.78	0.77	0.77	0.76	0.76	0.77	0.78	0.77
B CART	0.83	0.85	0.84	0.81	0.82	0.83	0.82	0.78	0.74	0.78	0.77	0.73	0.77	0.78
RF	0.88	0.88	0.87	0.84	0.84	**0.87**	0.86	0.83	0.78	0.78	0.79	0.79	**0.84**	0.80
SVM L	0.79	0.78	0.78	0.81	0.78	0.76	0.80	0.66	0.66	0.66	0.66	0.66	0.66	0.57
SVM P	0.85	0.84	0.85	0.87	0.87	**0.85**	0.85	0.86	0.84	0.84	0.84	0.84	**0.84**	0.83
SVM R	0.78	0.76	0.79	0.76	0.77	0.78	0.75	0.79	0.77	0.79	0.74	0.76	0.78	0.77
SGB	0.84	0.85	0.86	0.86	0.86	0.86	0.84	0.77	0.77	0.80	0.79	0.79	0.77	0.79
QDA	0.64	0.63	0.68	0.69	0.68	0.65	0.69	0.60	0.60	0.60	0.60	0.60	0.60	0.60
LDA	0.71	0.67	0.69	0.69	0.73	0.74	0.74	0.63	0.69	0.71	0.66	0.64	0.66	0.62

Realizado por: Robles, Alberto, 2022

Table 4. Optimization of the nonAPP - ALP classification algorithms taking into account Kendall's correlation cutoffs 1, 0.95, 0.90, 0.85, 0.80, 0.75.

Corr	Model (105/105)							Evaluation (45/45)						
	1	0.95	0.9	0.85	0.8	0.75	0.7	1	0.95	0.9	0.85	0.8	0.75	0.7
N	88	86	85	84	81	76	72	88	86	85	84	81	76	72
LR	0.73	0.76	0.77	0.75	0.77	0.75	0.77	0.80	0.81	0.78	0.78	0.79	0.81	0.81
NB	0.78	0.78	0.77	0.79	0.80	0.78	0.80	0.79	0.79	0.79	0.79	0.79	0.79	0.78
B CART	0.88	0.89	0.86	0.87	0.88	0.88	0.86	0.83	0.84	0.82	0.83	0.81	0.84	0.82
RF	0.89	0.90	0.89	0.89	0.91	0.89	**0.92**	0.88	0.87	0.87	0.83	0.88	0.88	**0.89**
SVM L	0.76	0.79	0.78	0.77	0.79	0.80	0.80	0.79	0.78	0.79	0.79	0.79	0.79	0.80
SVM P	0.90	0.90	0.89	0.89	0.90	0.89	**0.90**	0.88	0.88	0.88	0.88	0.88	0.90	**0.88**
SVM R	0.83	0.84	0.82	0.83	0.83	0.82	0.84	0.82	0.84	0.82	0.79	0.83	0.79	0.80
SGB	0.89	0.88	0.86	0.89	0.90	0.89	**0.89**	0.84	0.82	0.84	0.88	0.84	0.87	**0.88**
QDA	0.71	0.70	0.71	0.71	0.71	0.70	0.70	0.63	0.63	0.63	0.68	0.63	0.63	0.63
LDA	0.75	0.74	0.73	0.72	0.71	0.72	0.74	0.77	0.77	0.76	0.76	0.77	0.77	0.74

Realizado por: Robles, Alberto, 2022

Design of new peptide sequences after generating the 5000 sequences

With the modlamp package and testing them with the 5 developed classification algorithms, 221 peptides potential activity against the *Leishmania* parasite. In line with this, 74% of these molecules were predicted as antiparasitic according to *in silico* analysis by AMPfun tool [23]. Likewise, 100% of the designed molecules are not found within the LAMP2, APD3, DBAASP, DRAMP2 and Inverpep databases (Search date: March 25, 2022), demonstrating that all these sequences are new or discovered. A deeper view on frequency of the natural amino acids of the 221 sequences, it shows that Lysine (K) was the residues with the highest percentage of abundance (**Figure 4**). According to the recent study by Robles and collaborators, this amino acid cause ALPs to have the properties of positive charge to interact with and destabilize the parasite membrane. On the other hand, using the *in silico* PEP2D tool, it is shown that more than 50% of the designed antileishmania structures are alpha helical in shape. Martins et. Al., Rodriguez Vera and Cobb suggest that this molecular architecture facilitates the interaction and induction of membrane damage [50-52].

Approximately twice as many promising candidates have been determined with this research as have been discovered in the entire history of antileishmanial peptides. Future studies should perform experimental validation of these peptides to know their activity and potency, such as the inhibition concentrations. The toxicity and hemolytic effects should also be evaluated. These algorithms can accelerate the daunt process of finding an effective therapy against *Leishmania* parasites. Finally, the increase in the

databases should allow the elaboration of more specific models. For example, to differentiate between ALPs and non-ALPs or even to predict the IC50 value.

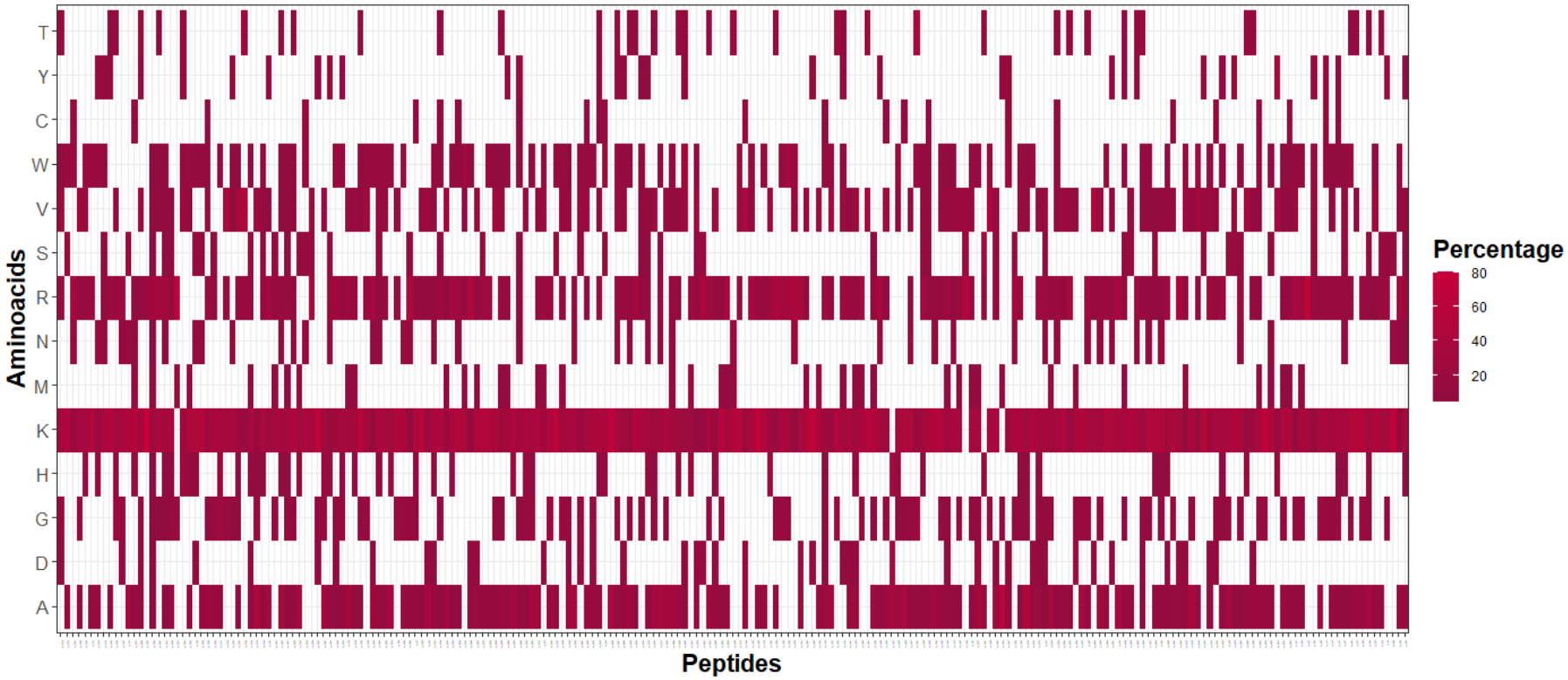

Figure 4. Analysis of amino acid percentage of potential antileishmanial peptides.
Realizado por: Robles, Alberto, 2022

CONCLUSION

In total, out of 5000 random sequences designed, 221 were established as promising anti-leishmanial peptides. None of these were found in the most comprehensive databases of antiparasitic and antimicrobial peptides such as LAMP2, APD3, DBAASP, DRAMP2 and Inverpep, thus establishing them as novel. The accuracy of the algorithms on the training data was RF - 92%, SVM P - 90%, SGB - 89% for the classification of non-APP - ALP and RF - 87%, SVM P - 85% for the prediction of AMP – ALP classes. While for the test data it was RF - 89%, SVM P - 88%, SGB - 88% for the non-APP – ALP classification and RF - 84%, SVM P - 84% for the prediction of the AMP - ALP classes. This is the first research at the artificial intelligence level that focused on predicting and discovering new anti-infective therapeutics targeting *Leishmania* parasites. Our results practically doubled the number of antileishmanial peptides that have been reported in literature and registered in current databases.

REFERENCES

1. WHO. Ending the neglect to attain the sustainable development goals: a road map for neglected tropical diseases 2021–2030: overview. Available from: https://apps.who.int/iris/handle/10665/332094.
2. Steverding D. The history of leishmaniasis. Parasites & vectors. 2017;10(1):1-10.
3. WHO. Leishmaniasis 2021 [cited 2022 13 enero]. Available from: https://www.who.int/news-room/fact-sheets/detail/leishmaniasis.
4. WHO. Leishmaniasis 2021 [cited 2022 13 enero 2022]. Available from: https://www.who.int/es/news-room/factheets/detail/leishmaniasis#:~:text=Hay%20tres%20formas%20principales%20de,picadura%20de%20fleb%C3%B3tomos%20hembra%20infectados.
5. WHO. Leishmaniasis 2022 [cited 2022 09 April]. Available from: https://www.paho.org/en/topics/leishmaniasis.
6. Roatt BM, de Oliveira Cardoso JM, De Brito RCF, Coura-Vital W, de Oliveira Aguiar-Soares RD, Reis AB. Recent advances and new strategies on leishmaniasis treatment. Applied Microbiology and Biotechnology. 2020;104(21):8965-77.
7. Pradhan S, Schwartz R, Patil A, Grabbe S, Goldust M. Treatment options for leishmaniasis. Clinical and experimental dermatology. 2022;47(3):516-21.
8. Sundar S, Chakravarty J, Meena LP. Leishmaniasis: treatment, drug resistance and emerging therapies. Expert Opinion on Orphan Drugs. 2019;7(1):1-10.
9. Oliveira SS, Ferreira CS, Branquinha MH, Santos AL, Chaud MV, Jain S, et al. Overcoming multi-resistant leishmania treatment by nanoencapsulation of potent antimicrobials. Journal of Chemical Technology & Biotechnology. 2021;96(8):2123-40.
10. Mookherjee N, Anderson MA, Haagsman HP, Davidson DJ. Antimicrobial host defence peptides: functions and clinical potential. Nature reviews Drug discovery. 2020;19(5):311-32.
11. Usmani SS, Bedi G, Samuel JS, Singh S, Kalra S, Kumar P, et al. THPdb: Database of FDA-approved peptide and protein therapeutics. PloS one. 2017;12(7):e0181748.
12. Robles-Loaiza AA, Pinos-Tamayo EA, Mendes B, Teixeira C, Alves C, Gomes P, et al. Peptides to Tackle Leishmaniasis: Current Status and Future Directions. International Journal of Molecular Sciences. 2021;22(9):4400. PubMed PMID: doi:10.3390/ijms22094400.
13. Capecchi A, Cai X, Personne H, Köhler T, van Delden C, Reymond J-L. Machine learning designs non-hemolytic antimicrobial peptides. Chemical Science. 2021;12(26):9221-32.
14. Basith S, Manavalan B, Shin TH, Lee DY, Lee G. Evolution of machine learning algorithms in the prediction and design of anticancer peptides. Current Protein and Peptide Science. 2020;21(12):1242-50.
15. Robles-Loaiza AA, Pinos-Tamayo EA, Mendes B, Ortega-Pila JA, Proaño-Bolaños C, Plisson F, et al. Traditional and Computational Screening of Non-Toxic Peptides and Approaches to Improving Selectivity. Pharmaceuticals. 2022;15(3):323.
16. Aronica PG, Reid LM, Desai N, Li J, Fox SJ, Yadahalli S, et al. Computational methods and tools in antimicrobial peptide research. Journal of Chemical Information and Modeling. 2021;61(7):3172-96.
17. Chowdhury AS, Reehl SM, Kehn-Hall K, Bishop B, Webb-Robertson B-JM. Better understanding and prediction of antiviral peptides through primary and secondary structure feature importance. Scientific reports. 2020;10(1):1-8.
18. Zhang W, Xia E, Dai R, Tang W, Bin Y, Xia J. PredAPP: Predicting Anti-Parasitic Peptides with Undersampling and Ensemble Approaches. Interdisciplinary Sciences: Computational Life Sciences. 2021:1-11.
19. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Molecular diversity. 2021;25(3):1315-60.
20. Müller AT, Gabernet G, Hiss JA, Schneider G. modlAMP: Python for antimicrobial peptides. Bioinformatics. 2017;33(17):2753-5.
21. Cao D-S, Xiao N, Xu Q-S, Chen AF. Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions. Bioinformatics. 2015;31(2):279-81.
22. Lunardon N, Menardi G, Torelli N. ROSE: a package for binary imbalanced learning. R journal. 2014;6(1).
23. Chung C-R, Kuo T-R, Wu L-C, Lee T-Y, Horng J-T. Characterization and identification of antimicrobial peptides with different functional activities. Briefings in bioinformatics. 2020;21(3):1098-114.
24. Joseph S, Karnik S, Nilawe P, Jayaraman VK, Idicula-Thomas S. ClassAMP: a prediction tool for classification of antimicrobial peptides. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2012;9(5):1535-8.
25. Meher PK, Sahu TK, Saini V, Rao AR. Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. Scientific reports. 2017;7(1):1-12.

26.Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010;26(5):680-2.
27.Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-9.
28.Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics. 2001;17(3):282-3.
29.Li W, Jaroszewski L, Godzik A. Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics. 2002;18(1):77-82.
30.Plisson F, Ramírez-Sánchez O, Martínez-Hernández C. Machine learning-guided discovery and design of non-hemolytic peptides. Scientific Reports. 2020;10(1):16581. doi: 10.1038/s41598-020-73644-6.
31.Singh H, Singh S, Raghava GPS. Peptide secondary structure prediction using evolutionary information. BioRxiv. 2019:558791.
32.Oliveira AL. Biotechnology, big data and artificial intelligence. Biotechnology journal. 2019;14(8):1800613.
33.Lin T-T, Yang L-Y, Lu I-H, Cheng W-C, Hsu Z-R, Chen S-H, et al. AI4AMP: an Antimicrobial Peptide Predictor Using Physicochemical Property-Based Encoding Method and Deep Learning. Msystems. 2021;6(6):e00299-21.
34.Singh O, Hsu W-L, Su EC-Y. Co-AMPpred for in silico-aided predictions of antimicrobial peptides by integrating composition-based features. BMC bioinformatics. 2021;22(1):1-21.
35.Maćkiewicz A, Ratajczak W. Principal components analysis (PCA). Computers & Geosciences. 1993;19(3):303-42.
36.Giovati L, Ciociola T, Magliani W, Conti S. Antimicrobial peptides with antiprotozoal activity: current state and future perspectives. Future Science; 2018. p. 2569-72.
37.Liu S, Bao J, Lao X, Zheng H. Novel 3D structure based model for activity prediction and design of antimicrobial peptides. Scientific reports. 2018;8(1):1-12.
38.Yan J, Zhang B, Zhou M, Kwok HF, Siu SW. Multi-Branch-CNN: Classification of ion channel interacting peptides using multi-branch convolutional neural network. Computers in Biology and Medicine. 2022:105717.
39.Chadbourne FL, Raleigh C, Ali HZ, Denny PW, Cobb SL. Studies on the antileishmanial properties of the antimicrobial peptides temporin A, B and 1Sa. Journal of Peptide Science. 2011;17(11):751-5.
40.Abbassi F, Raja Z, Oury B, Gazanion E, Piesse C, Sereno D, et al. Antibacterial and leishmanicidal activities of temporin-SHd, a 17-residue long membrane-damaging peptide. Biochimie. 2013;95(2):388-99.
41.Chadbourne F. The design and Synthesis of peptide-inspired antileishmanial agents: Durham University; 2014.
42.Mendes B, Proaño-Bolaños C, Gadelha FR, Almeida JR, Miguel DC. Cruzioseptins, antibacterial peptides from Cruziohyla calcarifer skin, as promising leishmanicidal agents. Pathogens and Disease. 2020;78(6):ftaa053.
43.Kückelhaus SA, Leite JRS, Muniz-Junqueira MI, Sampaio RN, Bloch Jr C, Tosta CE. Antiplasmodial and antileishmanial activities of phylloseptin-1, an antimicrobial peptide from the skin secretion of Phyllomedusa azurea (Amphibia). Experimental parasitology. 2009;123(1):11-6.
44.Pinto EG, Pimenta DC, Antoniazzi MM, Jared C, Tempone AG. Antimicrobial peptides isolated from Phyllomedusa nordestina (Amphibia) alter the permeability of plasma membrane of Leishmania and Trypanosoma cruzi.
Experimental parasitology. 2013;135(4):655-60.
45.Pinto EG, Antoniazzi MM, Jared C. Antileishmanial and antitrypanosomal activity of the cutaneous secretion of Siphonops annulatus. Journal of Venomous Animals and Toxins including Tropical Diseases. 2015;20:1-8.
46.X Chaves R, V Quelemes P, M Leite L, SA Aquino D, V Amorim L, AF Rodrigues K, et al. Antileishmanial and immunomodulatory effects of Dermaseptin-01, a promising peptide against Leishmania amazonensis. Current Bioactive Compounds. 2017;13(4):305-11.
47.Pérez-Cordero JJ, Lozano JM, Cortés J, Delgado G. Leishmanicidal activity of synthetic antimicrobial peptides in an infection model with human dendritic cells. Peptides. 2011;32(4):683-90.
48.Kulkarni MM, Karafova A, Kamysz W, McGwire BS. Design of protease-resistant pexiganan enhances antileishmanial activity. Parasitology research. 2014;113(5):1971-6.
49.Alin A. Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics. 2010;2(3):370-4.
50.Martins DB, Vieira MR, Fadel V, Santana VAC, Guerra MER, Lima ML, et al. Membrane targeting peptides toward antileishmanial activity: design, structural determination and mechanism of interaction. Biochimica et Biophysica Acta (BBA)-General Subjects. 2017;1861(11):2861-71.
51.Rodríguez Vega CA. Síntesis y determinación estructural de péptidos derivados de Dermaseptina con actividad antileishmanial. Escuela de Química. 2011.
52.Cobb SL, Denny PW. Antimicrobial peptides for leishmaniasis. Current opinion in investigational drugs. 2010;11(8):868-75.

SUPPLEMENTARY DATA

Supplemental definitions:

- **Accuracy:** describe of performance across all classes.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** measures the accuracy of the model when it predicts the positive class.

$$Prec = \frac{TP}{TP + FP}$$

- **Recall:** measures the ability of the model to predict positive classes.

$$Recall = \frac{TP}{TP + FN}$$

- **Cohen's Kappa statistic:** compares the observed precision with the expected one.

$$KAPPA = \frac{observed\ Acc - expected\ Acc}{1 - expected\ Acc}$$

- **F1 – score:** defined as the harmonic mean between precision and recall.

$$F1 = 2 * \frac{Prec * Recall}{Prec + recall}$$

- **Matthew's correlation coefficient (MCC):** is a measure of the association between two binary variables.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

- **Area under the curve Receiver Operating Characteristic (AUC-ROC):** the performance of a classification model at all classification thresholds.

Table S1. Measures of central tendency, dispersion, asymmetry, kurtosis and normality of the 56 physical-chemical descriptors of antimicrobial peptides.

Variable	Mean	SD	Median	Min	Max	25th	75th	Skew	Kurtosis	Statistic	p - value	Normality
H_Eisenberg	0.09	0.29	0.10	-0.60	0.73	-0.13	0.28	0.03	-0.54	0.99	0.5662	YES
uH_Eisenberg	0.26	0.15	0.25	0.00	0.63	0.13	0.36	0.45	-0.70	0.96	2.E-04	NO
H_GRAVY	0.11	0.83	0.07	-2.46	1.96	-0.43	0.70	-0.01	-0.16	0.99	0.1771	YES
uH_GRAVY	0.80	0.52	0.70	0.00	2.05	0.37	1.19	0.54	-0.74	0.95	<0.001	NO
Z3_1	8.34	2.16	7.93	4.81	14.05	6.69	9.55	0.65	-0.27	0.95	1E-04	NO
Z3_2	5.62	2.57	5.22	2.02	20.55	4.15	6.41	3.11	14.49	0.74	<0.001	NO
Z3_3	3.69	1.61	3.77	0.30	8.75	2.47	4.81	0.16	-0.35	0.99	0.2095	YES
Z5_1	7.87	1.83	7.57	4.83	13.76	6.54	8.75	0.78	0.24	0.95	1.E-04	NO
Z5_2	4.65	1.49	4.41	2.07	12.48	3.81	5.37	2.22	8.57	0.83	<0.001	NO
Z5_3	3.40	1.33	3.43	0.11	7.85	2.38	4.28	0.36	0.24	0.98	0.0695	YES
Z5_4	1.90	0.88	1.76	0.73	6.06	1.33	2.29	1.93	5.45	0.84	<0.001	NO
Z5_5	1.19	0.60	1.03	0.17	2.67	0.65	1.77	0.28	-1.15	0.94	<0.001	NO
S_AASI	2.07	0.19	2.10	1.46	2.74	1.98	2.19	-0.27	1.88	0.95	<0.001	NO
uS_AASI	0.13	0.08	0.12	0.00	0.54	0.07	0.18	1.27	3.39	0.92	<0.001	NO
modlas_AHPRK	1.05	0.21	1.03	0.33	1.64	0.91	1.16	0.48	0.93	0.96	2E-04	NO
H_argos	0.10	0.30	0.05	-0.59	1.11	-0.09	0.23	0.73	0.75	0.95	1.E-04	NO
uH_argos	0.26	0.18	0.22	0.00	0.88	0.12	0.37	0.83	0.23	0.94	<0.001	NO
X_uilkiness	0.63	0.09	0.62	0.19	0.86	0.59	0.68	-1.51	6.42	0.88	<0.001	NO
u_uilkiness	0.08	0.06	0.06	0.00	0.27	0.04	0.10	1.32	1.56	0.88	<0.001	NO
charge_phys	0.11	0.10	0.11	-0.10	0.40	0.05	0.19	0.23	-0.12	0.99	0.4301	YES
charge_acid	0.13	0.10	0.13	-0.09	0.40	0.07	0.20	0.20	-0.22	0.99	0.2484	YES
Ez	19.70	1.02	19.53	17.07	23.14	19.07	20.11	0.80	1.01	0.95	1.E-04	NO
flexiility	0.57	0.07	0.57	0.35	0.84	0.52	0.61	0.15	1.07	0.98	0.0991	YES
u_flexiility	0.08	0.04	0.07	0.00	0.19	0.04	0.10	0.53	-0.52	0.96	3.E-04	NO
Grantham	85.53	12.15	85.03	34.97	125.64	79.05	92.92	-0.31	2.97	0.95	<0.001	NO
H_HoppWoods	-0.16	0.43	-0.17	-1.53	0.92	-0.40	0.11	-0.07	0.40	0.99	0.338	YES
uH.HoppWoods	0.44	0.25	0.41	0.00	1.21	0.22	0.61	0.54	-0.39	0.96	6E-04	NO
ISAECI	88.78	16.18	86.70	35.46	137.36	78.89	100.06	0.00	0.58	0.98	0.0199	NO
H_Janin	0.07	0.28	0.09	-0.79	0.67	-0.10	0.26	-0.39	0.00	0.99	0.1887	YES
uH_Janin	0.26	0.15	0.23	0.00	0.64	0.14	0.37	0.41	-0.59	0.97	0.002	NO
H_KyteDoolittle	0.21	0.28	0.20	-0.65	0.82	0.03	0.41	-0.03	-0.15	0.99	0.1686	YES
uH_KyteDoolittle	0.27	0.17	0.24	0.00	0.70	0.12	0.40	0.57	-0.69	0.94	<0.001	NO
F_Levitt	1.00	0.08	1.01	0.72	1.17	0.96	1.04	-0.83	1.61	0.96	1.E-04	NO
uF_Levitt	0.07	0.05	0.05	0.00	0.25	0.03	0.10	1.19	0.97	0.89	<0.001	NO

MSS_shape	18.29	1.92	18.61	8.27	22.09	17.55	19.49	-2.47	10.01	0.80	<0.001	NO
u_MSS_shape	1.36	0.92	1.13	0.00	5.34	0.64	1.84	1.20	1.78	0.91	<0.001	NO
MSW	-0.29	0.25	-0.33	-1.05	0.65	-0.46	-0.14	0.66	1.26	0.97	0.0012	NO
pepArc	1.11	0.15	1.12	0.42	1.50	1.04	1.18	-0.74	3.36	0.94	<0.001	NO
pepcats	1.78	0.32	1.81	0.86	2.71	1.59	2.00	-0.25	-0.02	0.99	0.3629	YES
polarity	0.39	0.07	0.39	0.21	0.61	0.35	0.43	-0.07	-0.03	0.99	0.4938	YES
u_polarity	0.08	0.05	0.07	0.00	0.19	0.05	0.11	0.58	-0.42	0.96	1.E-04	NO
PPCALI	-0.47	1.08	-0.50	-3.61	2.09	-1.17	0.30	-0.16	-0.16	0.99	0.7247	YES
refractivity	0.40	0.07	0.40	0.16	0.65	0.35	0.44	-0.13	1.42	0.98	0.0177	NO
u_refractivity	0.05	0.03	0.04	0.00	0.14	0.02	0.06	0.98	0.60	0.93	<0.001	NO
t_scale	-7.49	5.06	-7.71	-29.14	5.09	-9.93	-5.11	-0.62	3.20	0.93	<0.001	NO
TM_tend	-0.31	0.46	-0.38	-1.55	0.78	-0.60	0.00	0.32	-0.01	0.98	0.011	NO
u_TM_tend	0.46	0.29	0.37	0.00	1.22	0.21	0.70	0.51	-0.78	0.95	<0.001	NO
Length	27.75	10.01	28.00	11.00	48.00	19.00	34.00	0.17	-0.94	0.97	0.0013	NO
omanIndex	0.86	1.53	0.73	-2.27	4.95	-0.19	1.92	0.16	-0.53	0.99	0.4065	YES
Aromaticity	0.09	0.07	0.08	0.00	0.39	0.05	0.12	1.35	3.23	0.90	<0.001	NO
AliphaticIndex	87.99	49.48	84.33	0.00	227.50	50.44	115.30	0.56	-0.25	0.97	0.0014	NO
InstailityIndex	31.41	27.24	28.11	-25.35	111.23	13.09	46.48	0.66	0.23	0.97	0.0011	NO
NetCharge	2.89	2.91	2.87	-3.94	12.03	0.99	4.05	0.41	0.37	0.98	0.042	NO
MW	3034.90	1101.44	2922.34	1118.33	5594.49	2079.09	3949.43	0.25	-0.94	0.97	0.0019	NO
IsoelectricPoint	9.38	2.13	10.03	2.85	12.74	8.12	10.79	-1.07	0.81	0.91	<0.001	NO
HydrophoicRatio	0.43	0.13	0.42	0.00	0.68	0.35	0.52	-0.42	0.55	0.97	0.0034	NO

Realizado por: Robles, Alberto, 2022

Table S2. Measures of central tendency, dispersion, asymmetry, kurtosis and normality of the 56 physical-chemical descriptors of no antiparasitic peptides.

Variable	Mean	SD	Median	Min	Max	25th	75th	Skew	Kurtosis	Statistic	p - value	Normality
H_Eisenberg	0.07	0.23	0.07	-0.60	0.62	-0.08	0.20	-0.08	-0.09	0.07	<0.001	NO
uH_Eisenberg	0.24	0.14	0.21	0.03	0.70	0.12	0.33	0.68	-0.07	0.24	0.0228	NO
H_GRAVY	-0.08	0.70	-0.03	-1.80	1.78	-0.55	0.34	0.10	-0.18	-0.08	<0.001	NO
uH_GRAVY	0.74	0.47	0.58	0.09	2.20	0.38	1.07	0.77	-0.27	0.74	<0.001	NO
Z3_1	8.32	1.71	8.44	4.25	13.42	7.17	9.44	-0.03	-0.16	8.32	1E-04	NO
Z3_2	5.30	1.93	4.95	1.64	10.83	3.94	6.70	0.44	-0.25	5.30	0.0018	NO
Z3_3	3.27	1.52	2.89	0.67	8.75	2.17	3.95	1.06	1.03	3.27	<0.001	NO
Z5_1	7.98	1.48	8.00	4.58	12.04	7.02	8.90	0.10	-0.26	7.98	<0.001	NO
Z5_2	4.31	1.16	4.17	1.56	7.22	3.57	5.23	0.20	-0.46	4.31	0.0217	NO
Z5_3	3.00	1.15	2.81	1.25	7.38	2.14	3.68	1.17	1.64	3.00	<0.001	NO
Z5_4	2.07	0.78	2.06	0.70	4.95	1.47	2.60	0.58	0.55	2.07	2E-04	NO
Z5_5	1.11	0.60	0.92	0.23	3.54	0.67	1.43	1.25	1.65	1.11	<0.001	NO
S_AASI	2.12	0.18	2.15	1.50	2.58	2.04	2.22	-0.93	1.56	2.12	0.0082	NO
uS_AASI	0.15	0.10	0.13	0.01	0.62	0.08	0.20	1.57	3.30	0.15	<0.001	NO
modlas_AHPRK	1.08	0.19	1.07	0.50	1.50	0.96	1.21	-0.05	-0.03	1.08	<0.001	NO
H_argos	0.07	0.25	0.06	-0.43	0.84	-0.11	0.25	0.48	-0.08	0.07	0.0121	NO
uH_argos	0.25	0.17	0.20	0.02	0.69	0.11	0.36	0.74	-0.54	0.25	7E-04	NO
X_uilkiness	0.62	0.07	0.63	0.43	0.81	0.58	0.66	0.05	0.13	0.62	0.2156	YES
u_uilkiness	0.07	0.04	0.06	0.00	0.20	0.04	0.09	1.08	1.13	0.07	2E-04	NO
charge_phys	0.05	0.10	0.07	-0.23	0.33	0.00	0.13	-0.33	-0.11	0.05	<0.001	NO
charge_acid	0.07	0.11	0.08	-0.18	0.34	0.00	0.15	-0.21	-0.18	0.07	<0.001	NO
Ez	19.88	0.86	19.86	17.62	22.62	19.38	20.40	0.22	0.53	19.88	0.0533	YES
flexiility	0.59	0.06	0.58	0.41	0.76	0.55	0.63	0.06	0.24	0.59	<0.001	NO
u_flexiility	0.07	0.04	0.06	0.01	0.20	0.04	0.09	0.74	-0.09	0.07	0.0055	NO
Grantham	83.87	9.34	84.02	56.90	107.42	78.31	89.00	0.02	0.17	83.87	0.0026	NO
H_HoppWoods	-0.05	0.36	-0.06	-1.06	0.99	-0.27	0.14	0.24	0.57	-0.05	<0.001	NO
uH.HoppWoods	0.43	0.25	0.37	0.01	1.09	0.23	0.62	0.62	-0.45	0.43	0.0031	NO
ISAECI	85.38	12.23	84.35	53.85	127.42	77.63	93.80	0.39	0.74	85.38	0.0039	NO
H_Janin	0.02	0.21	0.03	-0.61	0.71	-0.11	0.15	-0.08	0.58	0.02	<0.001	NO
uH_Janin	0.24	0.15	0.20	0.01	0.67	0.12	0.33	0.68	-0.38	0.24	0.002	NO

H_KyteDoolittle	0.14	0.23	0.16	-0.42	0.76	-0.01	0.29	0.10	-0.20	0.14	<0.001	NO
uH_KyteDoolittle	0.25	0.16	0.19	0.03	0.75	0.13	0.35	0.79	-0.18	0.25	1E-04	NO
F_Levitt	1.01	0.08	1.01	0.81	1.20	0.96	1.06	-0.06	0.03	1.01	0.0101	NO
uF_Levitt	0.06	0.04	0.05	0.00	0.16	0.03	0.07	0.90	0.16	0.06	<0.001	NO
MSS_shape	18.30	1.64	18.49	13.76	21.57	17.44	19.65	-0.48	-0.31	18.30	0.2665	YES
u_MSS_shape	1.27	0.76	1.10	0.11	3.55	0.67	1.67	0.72	-0.19	1.27	<0.001	NO
MSW	-0.34	0.20	-0.37	-0.78	0.26	-0.48	-0.21	0.32	-0.09	-0.34	<0.001	NO
pepArc	1.13	0.13	1.13	0.68	1.58	1.04	1.21	0.16	0.97	1.13	<0.001	NO
pepcats	1.74	0.30	1.71	0.91	2.45	1.54	1.93	0.05	-0.50	1.74	2E-04	NO
polarity	0.42	0.07	0.42	0.21	0.63	0.38	0.46	-0.01	0.81	0.42	7E-04	NO
u_polarity	0.08	0.05	0.07	0.01	0.20	0.04	0.11	0.74	-0.42	0.08	0.0031	NO
PPCALI	-0.31	0.98	-0.33	-3.97	2.87	-0.92	0.29	-0.14	0.87	-0.31	<0.001	NO
refractivity	0.38	0.06	0.37	0.23	0.53	0.33	0.42	0.13	-0.48	0.38	<0.001	NO
u_refractivity	0.04	0.02	0.04	0.00	0.11	0.02	0.05	0.57	0.19	0.04	<0.001	NO
t_scale	-6.68	5.29	-6.80	-19.26	7.51	-10.35	-3.23	-0.05	-0.24	-6.68	0.0023	NO
TM_tend	-0.42	0.40	-0.42	-1.50	0.74	-0.66	-0.17	-0.02	0.25	-0.42	<0.001	NO
u_TM_tend	0.44	0.29	0.36	0.02	1.19	0.23	0.65	0.73	-0.61	0.44	<0.001	NO
Length	27.16	9.71	25.00	11.00	49.00	19.25	35.00	0.24	-0.89	27.16	<0.001	NO
omanIndex	1.07	1.29	0.97	-1.79	4.49	0.23	2.10	0.04	-0.53	1.07	<0.001	NO
Aromaticity	0.08	0.06	0.08	0.00	0.31	0.03	0.11	0.83	0.96	0.08	<0.001	NO
AliphaticIndex	84.75	44.06	83.54	0.00	195.00	48.75	112.40	0.29	-0.62	84.75	0.0184	NO
InstailityIndex	30.31	30.12	26.44	-25.88	128.13	11.13	47.14	0.71	0.66	30.31	<0.001	NO
NetCharge	1.38	2.79	1.62	-6.94	8.06	-0.01	3.10	-0.26	-0.08	1.38	<0.001	NO
MW	2972.59	1091.72	2764.70	1204.40	5360.14	1995.58	3779.88	0.35	-0.86	2972.59	<0.001	NO
IsoelectricPoint	8.42	2.61	9.42	3.29	12.41	6.75	10.69	-0.56	-1.01	8.42	<0.001	NO
HydrophoicRatio	0.40	0.11	0.40	0.11	0.67	0.32	0.46	0.14	-0.11	0.40	<0.001	NO

Realizado por: Robles, Alberto, 2022

Table S3. Measures of central tendency, dispersion, asymmetry, kurtosis and normality of the 56 physical-chemical descriptors of antileishmanial peptides.

Variable	Mean	SD	Median	Min	Max	25th	75th	Skew	Kurtosis	Statistic	p - value	Normality
H_Eisenberg	0.04	0.50	0.16	-1.72	0.72	-0.11	0.37	-1.71	3.31	0.84	<0.001	NO
uH_Eisenberg	0.37	0.14	0.39	0.06	0.69	0.25	0.46	0.14	-0.72	0.98	0.0228	NO
H_GRAVY	0.14	1.15	0.31	-3.05	2.02	-0.38	0.82	-0.93	0.60	0.93	<0.001	NO
uH_GRAVY	1.11	0.52	1.04	0.20	2.10	0.72	1.60	0.15	-1.20	0.95	<0.001	NO
Z3_1	9.52	2.42	9.72	1.52	14.40	8.42	10.96	-0.80	0.94	0.96	1E-04	NO
Z3_2	5.06	1.52	5.36	1.12	9.29	4.04	6.17	-0.19	-0.19	0.97	0.0018	NO
Z3_3	3.49	1.96	2.69	1.22	9.71	2.09	4.68	1.30	1.10	0.86	<0.001	NO
Z5_1	8.95	2.10	9.04	1.47	12.83	7.96	10.34	-0.86	1.54	0.95	<0.001	NO
Z5_2	4.40	1.07	4.51	0.85	7.40	3.53	5.08	-0.13	0.56	0.98	0.0217	NO
Z5_3	3.28	1.83	2.57	1.40	9.91	2.11	4.07	1.79	3.07	0.79	<0.001	NO
Z5_4	2.02	0.81	2.05	0.49	5.54	1.39	2.49	0.78	1.78	0.96	2E-04	NO
Z5_5	0.78	0.49	0.65	0.07	2.93	0.48	0.88	2.14	5.39	0.78	<0.001	NO
S_AASI	2.09	0.16	2.10	1.58	2.57	2.03	2.19	-0.34	0.91	0.98	0.0082	NO
uS_AASI	0.17	0.10	0.15	0.01	0.54	0.10	0.22	1.20	1.84	0.91	<0.001	NO
modlas_AHPRK	1.16	0.27	1.08	0.33	1.80	1.00	1.33	0.18	0.55	0.94	<0.001	NO
H_argos	0.24	0.33	0.24	-0.56	0.91	0.00	0.51	-0.29	-0.47	0.98	0.0121	NO
uH_argos	0.38	0.21	0.36	0.03	0.83	0.22	0.55	0.27	-0.90	0.97	7E-04	NO
X_uilkiness	0.67	0.07	0.67	0.52	0.86	0.64	0.71	0.10	-0.21	0.99	0.2156	YES
u_uilkiness	0.11	0.06	0.10	0.00	0.27	0.06	0.16	0.42	-0.58	0.96	2E-04	NO
charge_phys	0.19	0.17	0.17	-0.25	0.75	0.07	0.29	1.21	1.81	0.89	<0.001	NO
charge_acid	0.21	0.16	0.18	-0.25	0.75	0.10	0.29	1.19	2.20	0.90	<0.001	NO
Ez	19.51	1.27	19.45	16.47	22.29	18.80	20.20	0.05	0.02	0.98	0.0533	YES
flexiility	0.54	0.11	0.53	0.31	0.92	0.49	0.58	1.34	3.34	0.88	<0.001	NO
u_flexiility	0.10	0.04	0.10	0.02	0.20	0.06	0.13	0.23	-0.78	0.97	0.0055	NO
Grantham	95.18	13.83	92.48	60.82	125.64	84.71	107.42	0.21	-0.54	0.97	0.0026	NO
H_HoppWoods	-0.05	0.66	-0.15	-0.97	2.06	-0.40	0.22	1.46	2.57	0.86	<0.001	NO
uH.HoppWoods	0.65	0.28	0.72	0.09	1.31	0.41	0.83	0.08	-0.83	0.97	0.0031	NO
ISAECI	97.46	15.42	100.82	51.99	137.36	87.71	109.36	-0.47	0.02	0.97	0.0039	NO

H_Janin	-0.08	0.43	0.02	-1.26	0.81	-0.24	0.20	-0.80	0.45	0.94	<0.001	NO
uH_Janin	0.38	0.17	0.41	0.05	0.75	0.25	0.49	-0.09	-0.61	0.97	0.002	NO
H_KyteDoolittle	0.22	0.38	0.27	-0.82	0.85	0.04	0.44	-0.91	0.54	0.93	<0.001	NO
uH_KyteDoolittle	0.37	0.17	0.34	0.07	0.71	0.23	0.53	0.19	-1.15	0.95	1E-04	NO
F_Levitt	1.06	0.07	1.07	0.83	1.25	1.01	1.11	-0.56	0.51	0.98	0.0101	NO
uF_Levitt	0.08	0.06	0.07	0.01	0.30	0.04	0.11	1.03	0.63	0.90	<0.001	NO
MSS_shape	18.38	1.37	18.43	15.20	22.09	17.36	19.29	-0.09	-0.45	0.99	0.2665	YES
u_MSS_shape	1.86	1.14	1.58	0.22	5.36	1.06	2.59	1.07	0.63	0.91	<0.001	NO
MSW	-0.19	0.33	-0.29	-0.84	0.85	-0.41	-0.02	0.89	0.34	0.93	<0.001	NO
pepArc	1.20	0.18	1.16	0.88	1.80	1.07	1.30	1.09	1.32	0.92	<0.001	NO
pepcats	1.78	0.39	1.72	1.06	2.85	1.51	2.02	0.63	-0.09	0.96	2E-04	NO
polarity	0.39	0.09	0.38	0.22	0.68	0.33	0.44	0.66	0.70	0.96	7E-04	NO
u_polarity	0.12	0.05	0.12	0.01	0.22	0.07	0.16	-0.12	-0.96	0.97	0.0031	NO
PPCALI	-0.06	0.94	0.14	-3.05	2.09	-0.52	0.55	-0.74	0.39	0.95	<0.001	NO
refractivity	0.40	0.08	0.38	0.26	0.65	0.34	0.43	0.95	0.45	0.92	<0.001	NO
u_refractivity	0.06	0.03	0.05	0.01	0.19	0.03	0.07	1.57	2.96	0.87	<0.001	NO
t_scale	-4.96	5.54	-5.65	-16.37	14.50	-7.79	-2.12	0.58	0.58	0.97	0.0023	NO
TM_tend	-0.29	0.62	-0.20	-2.08	0.75	-0.60	0.15	-0.88	0.55	0.94	<0.001	NO
u_TM_tend	0.68	0.31	0.70	0.06	1.32	0.42	0.92	0.02	-1.24	0.95	<0.001	NO
Length	19.97	11.10	17.00	4.00	70.00	13.00	25.00	1.86	5.49	0.84	<0.001	NO
omanIndex	0.91	2.63	-0.04	-2.39	10.69	-0.54	1.66	2.05	4.67	0.78	<0.001	NO
Aromaticity	0.08	0.08	0.07	0.00	0.39	0.04	0.11	1.41	2.23	0.87	<0.001	NO
AliphaticIndex	112.48	52.94	112.75	0.00	232.31	78.92	146.43	-0.12	-0.62	0.98	0.0184	NO
InstailityIndex	32.25	70.11	16.82	-36.49	353.11	-0.53	33.95	3.19	10.51	0.59	<0.001	NO
NetCharge	3.27	2.75	2.99	-3.01	12.68	1.03	4.99	0.92	0.88	0.91	<0.001	NO
MW	2245.43	1199.05	1923.36	503.55	7584.78	1484.44	2667.14	1.81	4.90	0.85	<0.001	NO
IsoelectricPoint	10.47	1.62	10.71	3.94	13.05	10.03	11.36	-1.84	4.84	0.82	<0.001	NO
HydrophoicRatio	0.47	0.15	0.51	0.00	0.83	0.39	0.56	-0.82	0.80	0.94	<0.001	NO

Realizado por: Robles, Alberto, 2022

Table S4. Effect of homologous sequences making similarity cuts CD - HIT on AMP - ALP models

Identity cut-off (CD-HIT)	**Model (105/105)**						**Evaluation (45/45)**					
	1	0.90	0.80	0.70	0.60	0.50	1	0.90	0.80	0.70	0.60	0.50
Acc (LR)	0.76	0.75	0.72	0.79	0.80	0.85	0.67	0.71	0.77	0.74	0.70	0.72
Acc (NB)	0.79	0.73	0.76	0.74	0.76	0.72	0.77	0.77	0.74	0.69	0.62	0.78
Acc (BT)	0.83	0.77	0.79	0.82	0.86	0.84	0.78	0.81	0.84	0.76	0.80	0.81
Acc (RF)	0.88	0.82	0.82	0.88	0.89	0.85	0.83	0.84	0.81	0.80	0.81	0.84
Acc (SVM L)	0.79	0.68	0.74	0.72	0.80	0.86	0.66	0.69	0.73	0.69	0.73	0.70
Acc (SVM P)	0.85	0.81	0.86	0.83	0.89	0.86	0.86	0.86	0.82	0.84	0.77	0.86
Acc (SVM R)	0.78	0.76	0.78	0.80	0.77	0.75	0.79	0.71	0.77	0.73	0.71	0.73
Acc (SGB)	0.84	0.82	0.83	0.82	0.85	0.87	0.77	0.84	0.76	0.76	0.81	0.80
Acc (Step QDA)	0.64	0.63	0.61	0.68	0.72	0.59	0.60	0.60	0.66	0.63	0.56	0.57
Acc (LDA)	0.71	0.73	0.67	0.73	0.78	0.80	0.63	0.64	0.76	0.74	0.76	0.68

Realizado por: Robles, Alberto, 2022

Table S5. Effect of homologous sequences making similarity cuts CD - HIT on non-APP - ALP models

Identity cut-off (CD-HIT)	**Model (105/105)**						**Evaluation (45/45)**					
	1	0.90	0.80	0.70	0.60	0.50	1	0.90	0.80	0.70	0.60	0.50
Acc (LR)	0.73	0.71	0.71	0.75	0.78	0.79	0.80	0.81	0.77	0.72	0.80	0.82
Acc (NB)	0.78	0.82	0.76	0.80	0.80	0.75	0.79	0.74	0.78	0.74	0.71	0.66
Acc (BT)	0.88	0.84	0.86	0.85	0.88	0.84	0.83	0.82	0.81	0.81	0.89	0.82
Acc (RF)	0.89	0.87	0.89	0.87	0.90	0.87	0.88	0.83	0.88	0.87	0.90	0.88
Acc (SVM L)	0.76	0.74	0.75	0.76	0.81	0.74	0.79	0.74	0.76	0.74	0.80	0.74
Acc (SVM P)	0.90	0.89	0.86	0.87	0.90	0.89	0.88	0.89	0.89	0.87	0.91	0.91
Acc (SVM R)	0.83	0.82	0.81	0.81	0.80	0.79	0.82	0.72	0.84	0.74	0.83	0.79
Acc (SGB)	0.89	0.89	0.85	0.88	0.88	0.86	0.84	0.88	0.83	0.87	0.83	0.87
Acc (Step QDA)	0.71	0.73	0.70	0.75	0.69	0.67	0.63	0.74	0.80	0.76	0.74	0.63
Acc (LDA)	0.75	0.75	0.75	0.73	0.74	0.78	0.77	0.76	0.74	0.79	0.81	0.80

Realizado por: Robles, Alberto, 2022

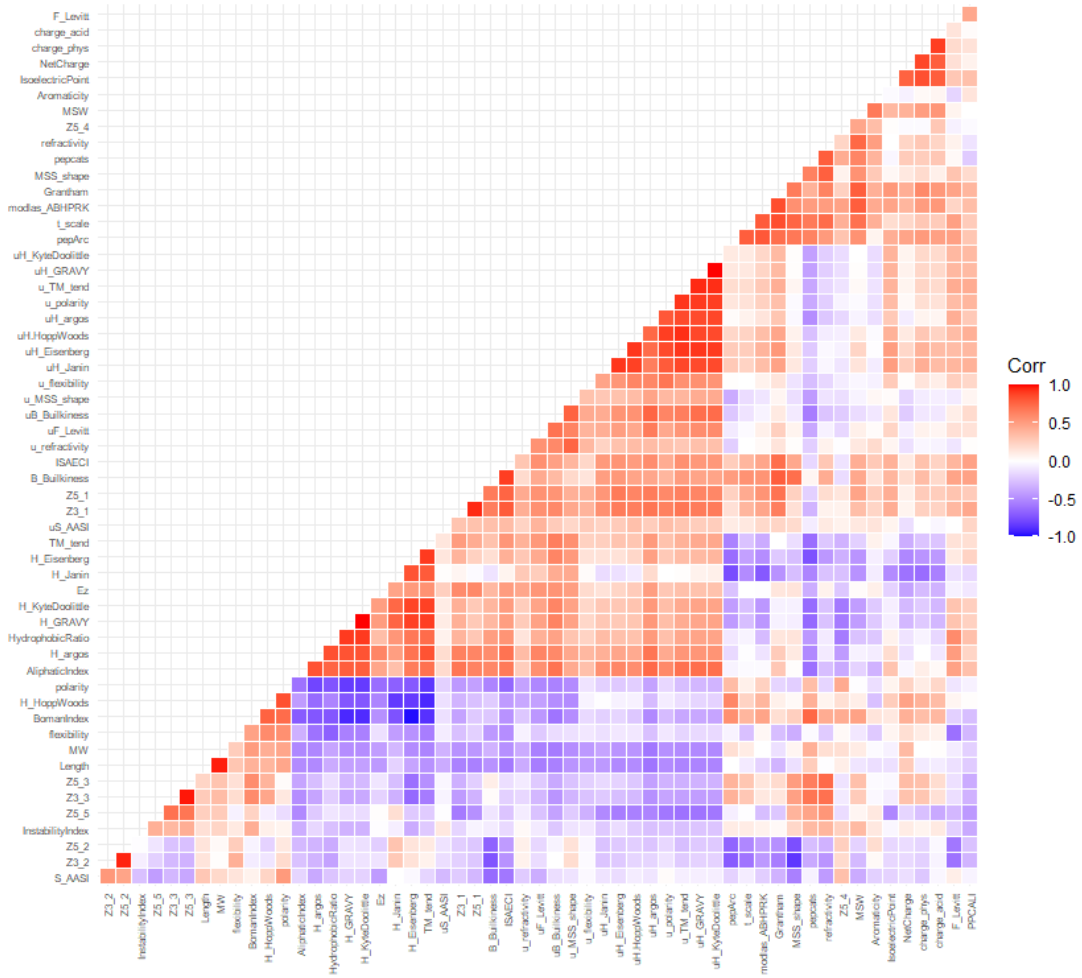

Figure S1. Pearson correlation matrix of the 56 physicochemical descriptors of antimicrobial peptides

Realizado por: Robles, Alberto, 2022

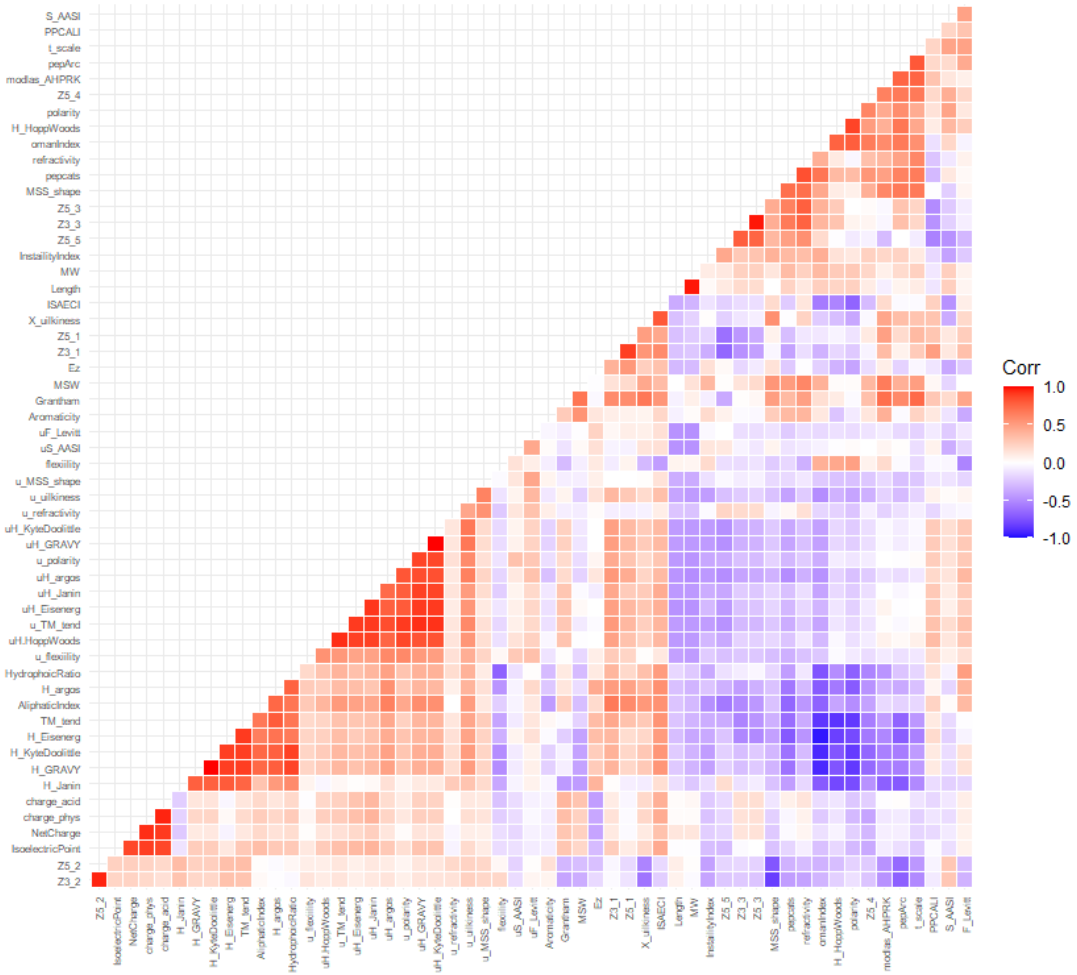

Figure S2. Pearson correlation matrix of the 56 physicochemical descriptors of non-antiparasitic peptides

Realizado por: Robles, Alberto, 2022

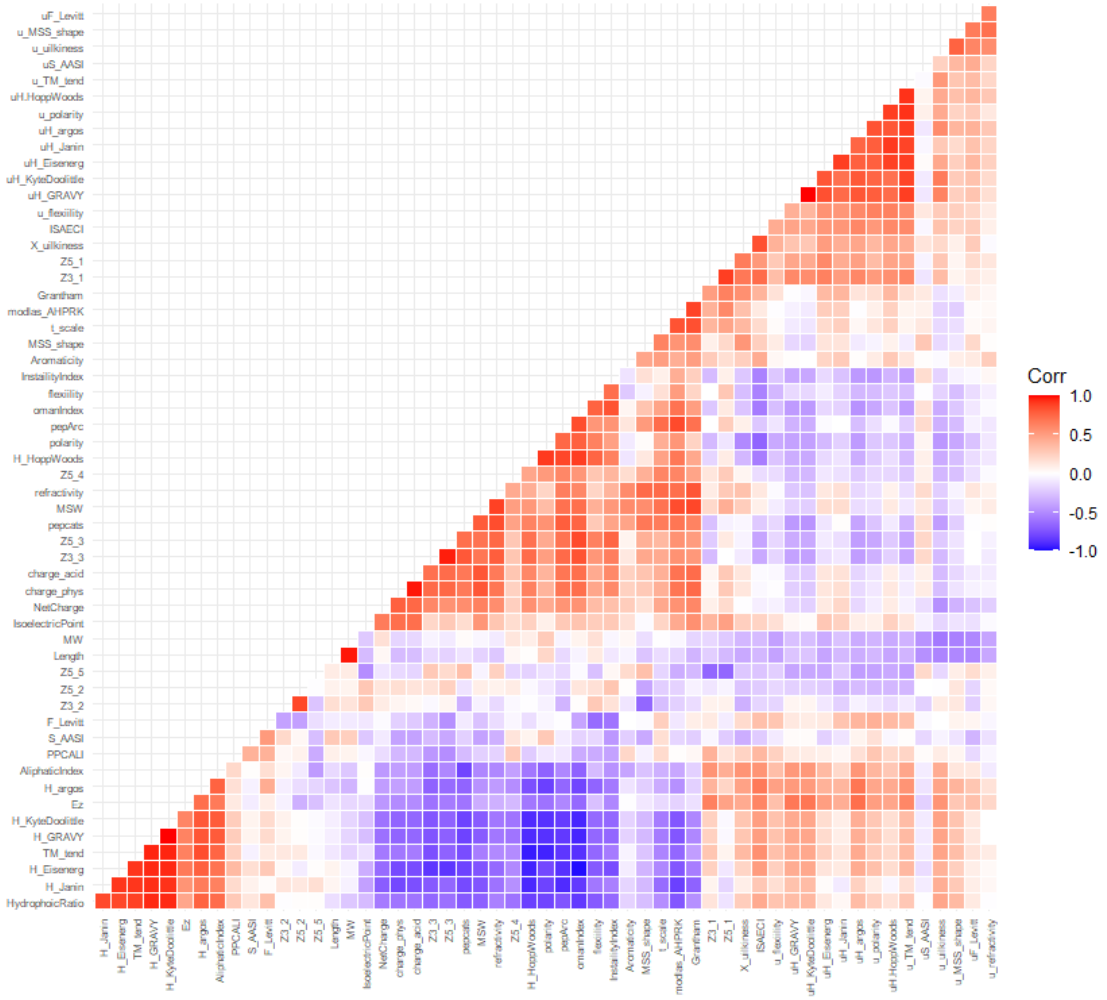

Figure S3. Pearson correlation matrix of the 56 physicochemical descriptors of antileishmanial peptides

Realizado por: Robles, Alberto, 2022

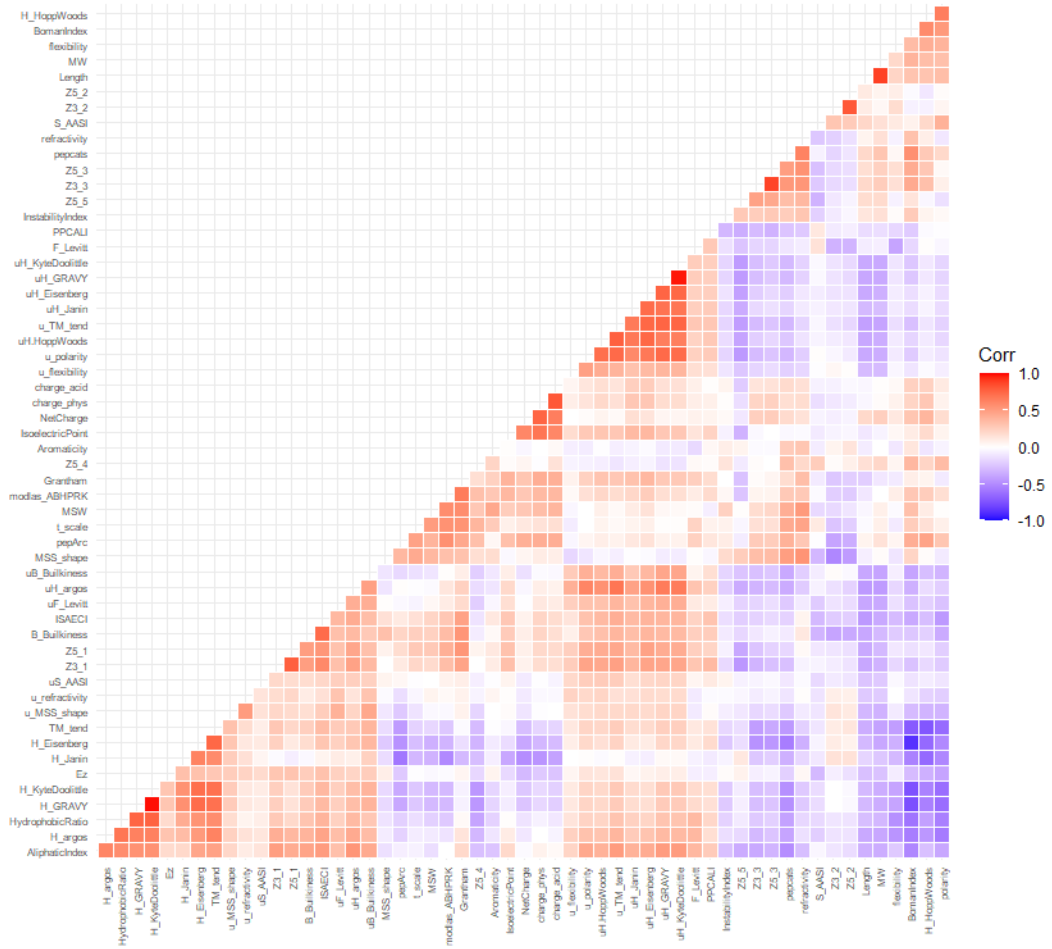

Figure S4. Kendall correlation matrix of the 56 physicochemical descriptors of antimicrobial peptides

Realizado por: Robles, Alberto, 2022

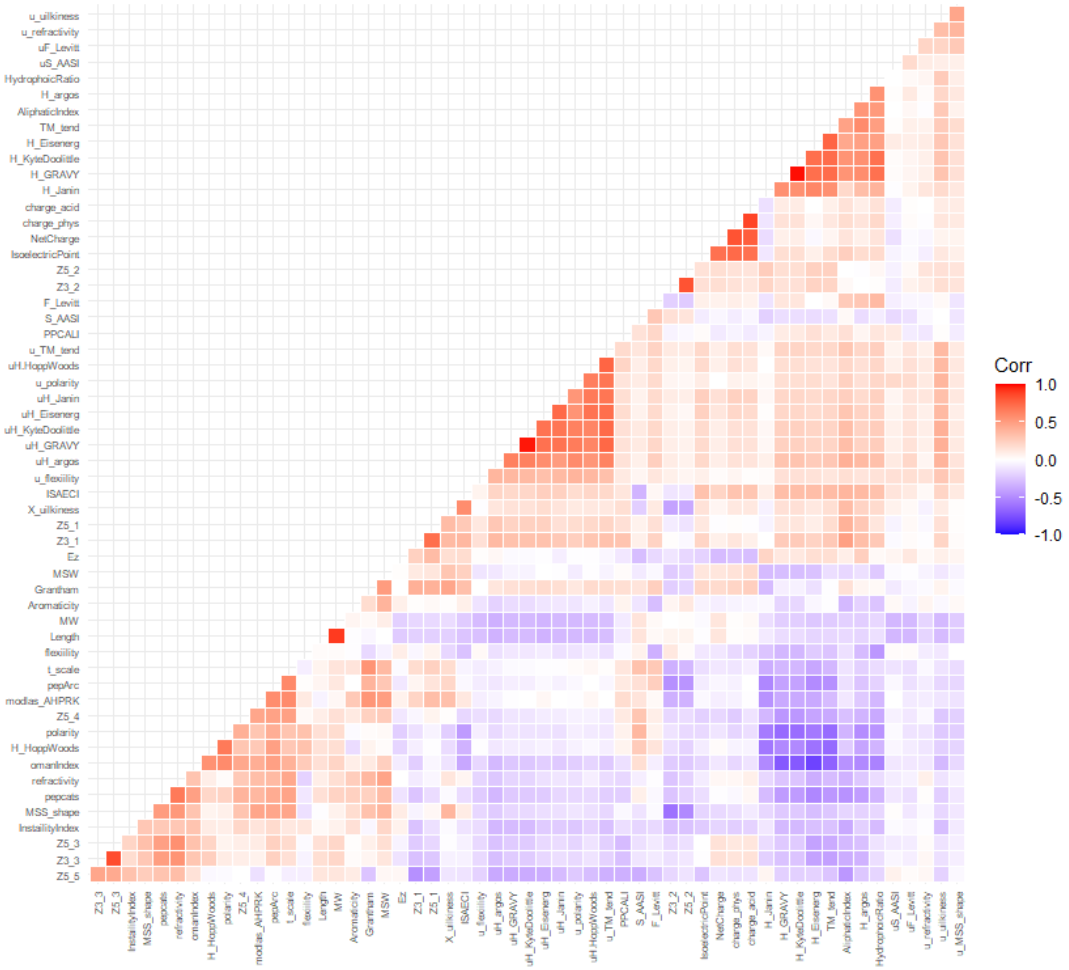

Figure S5. Kendall correlation matrix of the 56 physicochemical descriptors of antimicrobial peptides

Realizado por: Robles, Alberto, 2022

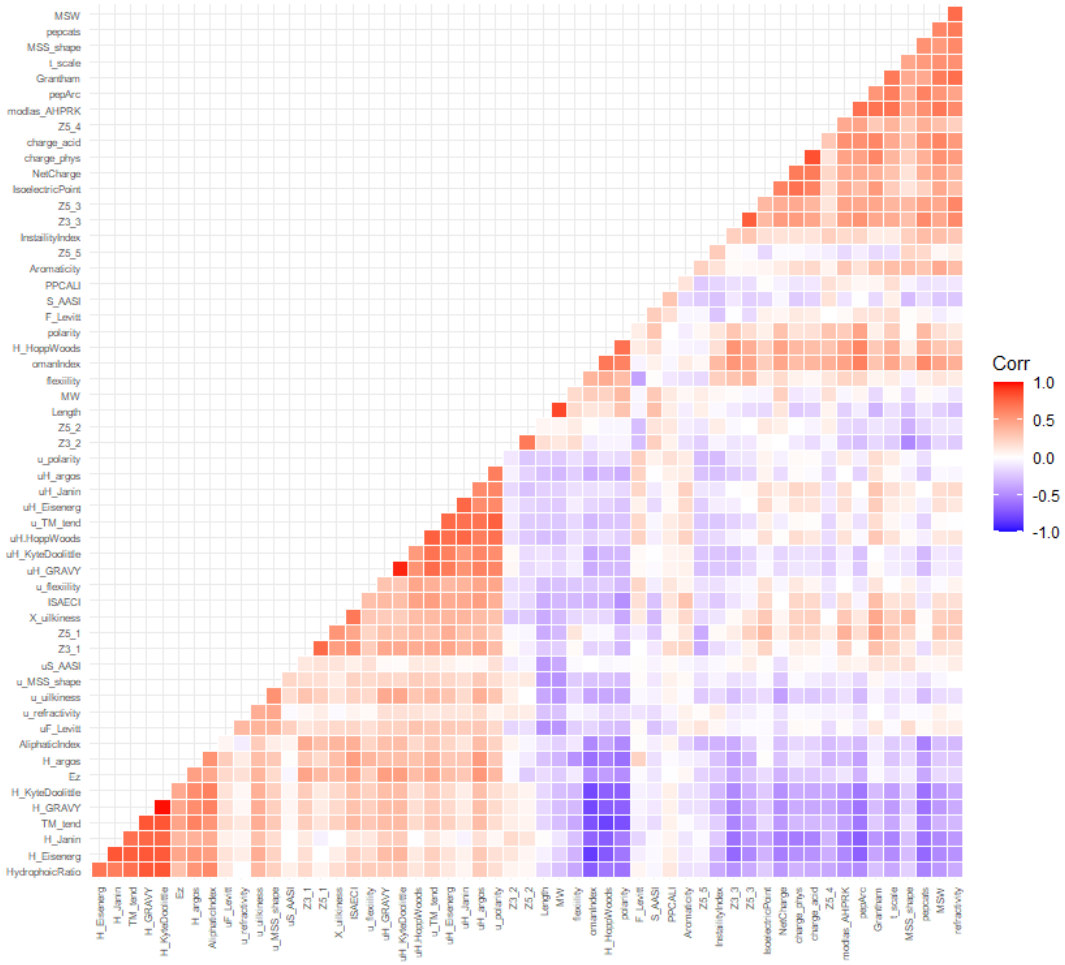

Figure S6. Kendall correlation matrix of the 56 physicochemical descriptors of antileishmanial peptides

Realizado por: Robles, Alberto, 2022

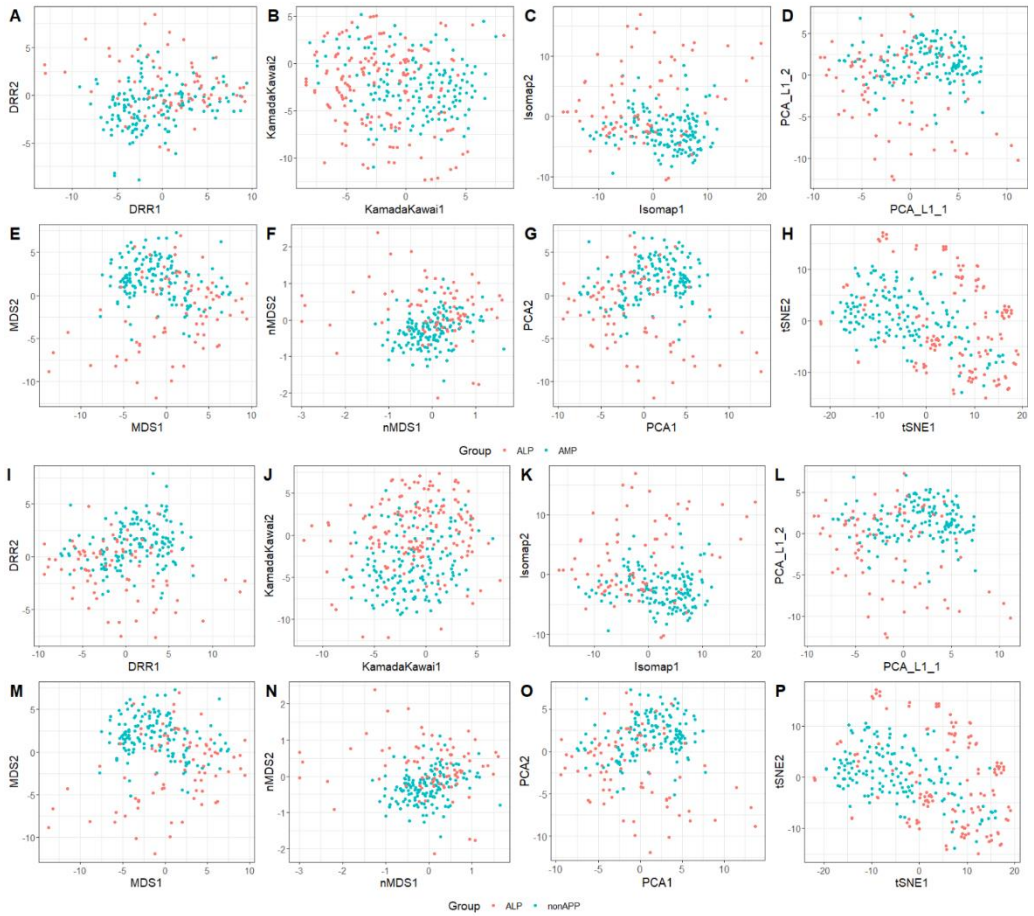

Figure S7. Graphic representation of component 1 and component 2 of the dimensionality reduction techniques: DRR, KamadaKawai, Isomap, PCA_L1, MDS, nMDS, PCA, tSNE.

Realizado por: Robles, Alberto, 2022