v. 5, n. 6, 2025

International Journal of Health Science

A SIMPLE UNSUPERVISED STACKING APPROACH TO IDENTIFY IN SILICO PEPTIDES WITH POTENTIAL ANTI-SARS-COV-2 ACTIVITY

Acceptance date: 17/01/2025

Raul Munoz-Hernandez
Departamento de Ciencias de la Computación, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada, México

Carlos A. Brizuela
Departamento de Ciencias de la Computación, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada, México

Abstract: The SARS-CoV-2 coronavirus, responsible for COVID-19 disease, has caused as of early February 2024, over 700 million cases worldwide and more than 6.9 million deaths. Great collaborative efforts have been made in the development of vaccines and therapeutic agents against SARS-CoV-2; however, little has been described regarding antiviral peptides (AVPs) despite their potential as antimicrobial agents. Interestingly, there are AVPs that have demonstrated prophylactic and therapeutic effects against coronaviruses. In this direction, targeting the Spike glycoprotein, some SARS-CoV-2 peptide inhibitors, like EK1, has been designed. In this work, Anti--SARS-CoV-2 peptides were gathered from the literature and the DRAMP database, and their molecular descriptors were computed by using the StarPep Toolbox. The same set of selected peptides was fed to AMPDiscover, a predictor for antimicrobial activity. The predictor is based on RF-models and predicts the following AMP activities: antibacterial, antifungal, antiparasitic and antiviral. The prediction of AMPDiscover is considered as a meta-descriptor to be used in a downstream unsupervised learning process. We use both representations and compared them in their ability to cluster similar peptides with respect to their activities by using the k-means algorithm. Preliminary results obtained through k-means clustering and Principal Component Analysis (PCA) techniques indicate an apparent grouping pattern of the tested Anti-SARS-CoV2 peptides in both cases. That is, it is possible to identify peptides with potential Anti-SARS-CoV-2 activity through specific clusters. In this task, the representation generated by the AMPDiscover shows better grouping capability than that generated by Starpep Toolbox.
Keywords: SARS-CoV-2, AVPs, k-means, unsupervised stacking.

INTRODUCTION

As of 1 November 2020, nearly 46 million cases of illness caused by COVID-19, and 1.2 million deaths were reported globally (World Health Organization, 2020). By early February 2024 over 700 million infection cases worldwide, and more than 6.9 million deaths, were documented[1]. Significant collaborative efforts have been made in the development of vaccines and therapeutic agents against SARS-CoV-2.

Many existing drugs with antiviral activity, speculated to be effective against COVID-19, have been rapidly brought into clinical trials. Among them are the antiparasitic drug ivermectin, the HIV protease inhibitor nelfinavir, and the anti-inflammatory drug cepharanthine (Caly et al., 2020; Ohashi et al., 2020). However, only dexamethasone and remdesivir have shown promising results (Ledford et al., 2020; Grein et al., 2020). The world is continuously searching for prophylactic compounds against SARS-CoV-2 and its new variants; however, little has been described regarding antiviral peptides (AVPs).

Antimicrobial peptides (AMPs) are part of our immune system and play an important role in combating infection-causing agents, such as bacteria, fungi, parasites, viruses and tumor cells (Wang *et al*., 2019). However, their mechanism of action in not fully understood. It is generally accepted that their antimicrobial ability is related to their ability to interact with the membrane of the invaders (Diamond *et al*., 2009; Zhang & Gallo, 2016; Zhang *et al*., 2021).

SARS-Cov-2 is an enveloped virus with a spike-shaped glycoprotein "Spike" (S). This mediates the entry of the virus through the ACE-2 (Angiotensin Converting Enzyme -2) cellular receptor (Zhou *et al*., 2020). The S1 subunit of the S protein is responsible for binding the ACE-2 site on the cell (Xia *et al*., 2020), and it shows 10 to 20 times higher affinity for the ACE-2 binding site compared to SARS-CoV, which explains a higher transmis-

sion and infection rate (Wrapp *et al*., 2020).

Subsequently, a serine protease called TMPRS22, is required to initiate the proteolytic cleavage of the S protein at the S1/S2 and S2' sites (Hoffmann *et al*., 2020). After cleavage, the heptapeptides HR1 and HR2, regions of the S2 subunit of the S protein, form the central 6-helix bundle (6HB) (Liu *et al*., 2004). The formation of the 6HB is critical in facilitating the process of viral membrane fusion, and allowing the virus to enter the cell via endocytosis (Ling *et al* 2020). The serine protease TMPRS22, then, can be a molecular target to be inhibited by an AVP.

SARS-CoV2 enters target cells through an endosomal pathway (Wang *et al*., 2008). Subsequently, during the late endosomal stage, the acidification of the endosome induces membrane fusion between the virus and the endosome. This could lead to the uncoating of the virus and the release of its RNA initiating the virus replication and infection process, as it has been demonstrated to occur in SARS--CoV (Du et al., 2009). However, it has been found that the mouse b-defensine-4 derived P9 peptide inhibits late endosomal acidification and thus preventing viral RNA release (Zhao et al., 2016).

There are AVPs that have demonstrated prophylactic and therapeutic effects against SARS-CoV and SARS-CoV-2 with several mechanisms of action involved. Some strategies, including "in silico" approaches have been tested to search for therapeutic agents (Essa *et al.,* 2022; Mahendran *et al*., 2020; Rani *et al.,* 2022). Some of the most important ones are described below.

APPROACHES TARGETING THE VIRAL ENVELOPE

Mucroporin-M is a peptide designed with four residue mutations, analogous to Mucroporin, isolated from the venom of the scorpion Lychas macronatus (Li et al., 2011). Recently, Mucroporin-M1 were studied using molecular docking and molecular dynamics with HR1 domain target of SARS-CoV-2, indicating that Mucroporin-M1 may have potential antiviral activity (Souza et al., 2023). In addition, molecular docking simulations were used to evaluate the actions of Mucroporin against ACE-2 receptors revealing for the first time that Mucroporin is functional inhibitor of ACE-2; so, it can be used as potential inhibitor to the ACE-2 receptor of SARS-CoV-2 (Fakih et al., 2022).

APPROACHES TARGETING THE SPIKE GLYCOPROTEIN

EK1 is a pan-CoV inhibitor designed by Xia and colleagues. EK1 acts by blocking the HR1 domain to disrupt the formation of the 6HB core, which leads to the inhibition of viral fusion with the host cell (Xia et al., 2020). Recently, a highly potent variant lipo-peptide, EK1C4, was generated by conjugating the C-terminal end of EK1 with a cholesterol moiety using glycine/serine linkers and a polyethylene glycol (PEG) spacer (Xia et al., 2019).

Cao et al. (2020) designed miniproteins using ACE2 as a scaffold and specific motifs from the RBD site in the S1 subunit of the spike protein of SARS-CoV-2. These proteins strongly adhere to the spike protein, preventing its attachment to ACE2. The most effective designs AHB1 & AHB2 (sequences derived from ACE), and LCB1 & LCB3 (de novo design), demonstrate exceptionally strong binding and effectively impede SARS-CoV-2 infection in mammalian Vero E6 cells.

APPROACHES INHIBITING THE ENDOSOMAL ACIDIFICATION

It was found that P9 binds to the S2 subunit of MERS-CoV and remain localized near the virus without inhibiting its entry via endocytosis (Zhao et al., 2016). In the endosomes, the polycationic nature of P9 induces a basic microenvironment that prevents the acidification of the late endosome and inhibits the activation of viral proteins that would initiate the fusion of the virus-cell membrane, uncovering the virus and releasing the viral RNA. P9 has demonstrated prophylactic and therapeutic activity "in vivo" against SARS-CoV.

To enhance the effectiveness of P9, Zhao et al (2020) introduced P9R. This newly developed peptide demonstrated robust antiviral properties against pH-dependent viruses, such as coronaviruses. Recently, a dual-functional cross-linking peptide known as 8P9R, derived from P9 and P9R, has been documented to impede virus entry via both the TMPRSS2 surface pathway and the endocytic pathway (Zhao et al., 2021).

APPROACHES PROTECTING CELLULAR RECEPTORS

It has been found that HD5, a lectin similar to human Defensin-5 (HD5), can interact with glycosylated proteins and lipid components (Wang, et al., 2020). As HD5 has much higher affinity to ACE2 receptors in comparison with that of SARS-CoV-2, competitively blocks ACE2 receptors on host cells. In addition, as it binds glycosylated Corona S1 protein, it prevents the virus invasion (Niv, 2020).

APPROACHES ACTING AS VIRAL IMMUNOMODULATORS

The cyclic peptide RTD-1 from Rhesus macaque leukocytes has been reported to reduce the pathogenicity of SARS-CoV infection in mice, despite not directly inhibiting the virus or interacting with host cell receptors (Wohlford--Lenane et al., 2009). Instead, RTD-1 was observed to modulate the host immune response. These findings suggest that RTD-1 acts as an immunomodulator via cytokine response.

As we could see in this brief literature review, the AVPs are structurally and functionally versatile due to their simple primary structure, and they can serve as molecular templates for the rapid generation of therapeutic agents for COVID-19 or emerging threats in the future. So, the purpose of this work is to discover, *in silico*, novel Anti-SARS-CoV-2 candidates from known AVPs. To this aim, AMPs with experimentally proven Anti-SARS-CoV-2 activity were retrieved from the literature (and their respective mechanisms of action, if available), their Molecular Descriptors calculated using StarPep (Aguilera et al, 2023), and their antimicrobial activities predicted using AMPDiscover (Pinacho-Castellanos et al, 2021), the predicted activities were used as a high-level molecular descriptor.

MATERIAL & METHODS

This section describes the process by which antiviral peptides with proven Anti-SARS-CoV2 activity were selected. Initially, a search was conducted in databases to identify peptides with documented Anti-SARS-CoV-2 activity. Among the information gathered, the DRAMP database was chosen to extract the sequences of the peptides. This choice was made because the database is updated and contains the majority of peptides of interest previously identified in the literature or other databases.

From the DRAMP database, an Anti-SARS-CoV-2 query was performed. The search produced 102 entries. As the peptides identified as DRAMP 29233 to DRAMP 29241 contained "x" in their sequences (amino acid not yet determined), they were removed from the final selection. The remaining 93 peptide sequences were selected for downstream process (Supplementary Material: S1).

From StarPep Toolbox a set of antiviral peptides were selected by applying the following steps: In "Query" window containing "All peptides" select "+", followed by "Add Target", and then "Virus" carpet followed by "OK". Following the steps detailed in Supplementary Material S2, a total of 3,062 antiviral peptides were obtained from Starpep Toolbox (Aguilera et al., 2023).

COMPUTING ANTI-VIRAL MOLECULAR DESCRIPTORS IN STARPEP

A total of 1037 molecular descriptors were calculated for each peptide by using the Starpep software (Aguilera et al, 2023). For details on the configuration used to compute the descriptors see Supplementary Material S2.

COMPUTING DESCRIPTORS FROM ACTIVITIES PREDICTION IN AMPDISCOVER

AMPDiscover is a computational platform to facilitate the exploration and analysis of antimicrobial peptides (AMPs), including bacteria, fungi, parasites, and viruses. It employs innovative methodologies to address inherent limitations in machine learning existing approaches, such as inadequate representation and redundancy in training datasets. By utilizing advanced modeling techniques, AMP Discover enables the accurate identification and classification of AMPs according to their antimicrobial properties. These models undergo rigorous validation processes, ensuring their reliability and effectiveness. AMP Discover offers open access to its models and tools at https://biocom-ampdiscover.cicese.mx/4.

These peptides along with the 93 selected from DRAMP were used as input queries for the AMPDiscover. AMPDiscover employs machine learning models based on Random Forest and ProtDCal descriptors (Romero-Molina, et al, 2019) to predict various antimicrobial activities of peptides. This platform provides predictions not only on the general antimicrobial activity of peptides but also on specific activities against bacteria, fungi, parasites, and viruses. AMPDiscover predicts the probability that a peptide exhibits a particular antimicrobial activity. These output values from AMPDiscover serve as indicators of confidence in the peptide's ability to combat specific microorganisms and are used as a high-level molecular descriptor. For details of these descriptors see Supplementary Material S2.

PHYSICOCHEMICAL DESCRIPTORS VS. PREDICTION OF BIOLOGICAL ACTIVITIES

One question we want to address is which representation for the peptides provide more information; the physicochemical properties--based descriptors as calculated by the StarPep software or the vector of predictions-based descriptors produced by AMPDiscover. A more informative representation will be the one that better separates or discriminates the biological information. To this aim, as a proof of concept, a simple clustering algorithm (k-means) and a dimensionality reduction algorithm (PCA) are used to study the representativeness of both approaches. We use the standard implementation in R for both algorithms (see https://rpubs.com/).

RESULTS & DISCUSSION

MOLECULAR DESCRIPTORS

Once the molecular descriptors of the selected peptides were calculated by StarPep, the optimal number of clusters to use with the k-means algorithm was determined with the elbow method (sum of squares), resulting in 4 clusters (Supplementary Material S9). To visualize them, the k-means algorithm and the PCA method was employed (Figure 1). Analyzing the resulting clustering, we can see that cluster No. 2 contained 50.5% of the known anti-SARS-CoV2. Additionally, 1509 new peptides were identified, belonging to the same cluster (1556 antivirals – 47 anti--SARS-CoV-2; Supplementary Material S10). However, this cluster represents 49.3% of the total peptides, so a biased selection of cluster membership for the Anti-SARS-CoV2 peptides cannot be ruled out (Table 1).

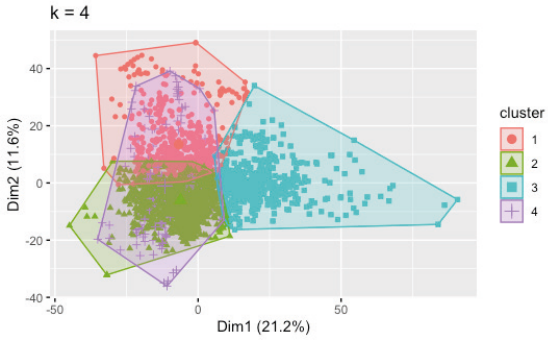


Fig. 1. PCA plot using the k-means algorithm at a level of k=4.

Cluster No. 2 contains 50.5% of the known Anti-SARS-CoV2.

k = 4	Total (%)	Anti-SARS-CoV2 (%)
Cluster 1	22,2	22,6
Cluster 2	49,3	50,5
Cluster 3	23,8	26,9
Cluster 4	4,7	0,0

Table 1. Comparative results (Total Peptides vs Anti-SARS-CoV-2 Peptides) when using the k-means clustering algorithm with k=4.

In order to mitigate the effect of randomness in clustering, the value of k was increased to 8 and then to 16. The k-means algorithm was applied for grouping, and PCA was used for visualization. The results for k=8 are shown in Figure 2. It was observed that when transitioning from k=4 to k=8, a cluster (No. 2) was obtained, grouping 59.1% of the included Anti-SARS-CoV2 in the run. Additionally, a total of 632 potential candidate peptides were identified, grouped in the same cluster (687 Antivirals – 55 Anti-SARS-CoV-2; Supplementary Material S11). It was also noted that this cluster represents 21.8% of the total peptides, thereby ruling out the possibility that these results are purely due to chance (Table 2).

In the case of k=16, it was observed that the majority of the known Anti-SARS-CoV2 were grouped in cluster No. 14 (41.9%), as depicted in Figure 3. This cluster also includes 456 candidate peptides (495 Antivirals – 39 Anti--SARS-CoV-2), and it represents only 15.7% of the total peptides studied (Supplementary Material: S12). Hence, it is also discarded that these clustering results are due to chance (Table 3). A summary of the results obtained for the Molecular Descriptors of the studied peptides is presented in Table 4.

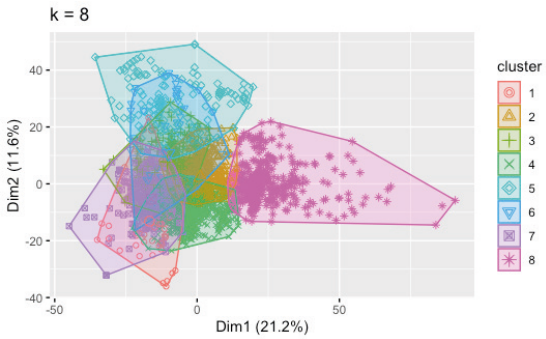


Fig. 2. PCA plot using the k-means algorithm at a level of k=8.

Cluster No. 2 contains 59.1% of the known Anti-SARS-CoV2.

Level of k = 8	Total (%)	Anti-SARS-CoV2 (%)
Cluster 1	1,8	0,0
Cluster 2	21,8	**59,1**
Cluster 3	12,7	7,5
Cluster 4	**23,6**	12,9
Cluster 5	5,1	2,2
Cluster 6	2,5	0,0
Cluster 7	13,0	10,8
Cluster 8	19,5	7,5

Table 2. Comparative results (Total Peptides vs Anti-SARS-CoV-2 Peptides) when using the k-means clustering algorithm with k=8.

BIOLOGICAL ACTIVITY PREDICTORS

Similar to what we did with the peptides represented by the physicochemical descriptors generated by Starpep, a clustering of peptides, applied to the feature vector composed of the predictions of AMPDiscover, was applied. This can be considered as an "unsupervised stacking" approach. Prior to using the k-means algorithm for clustering, the optimal number of clusters was calculated, resulting in 4 clusters (Elbow method). This result is similar to that obtained for the case of the physicochemical descriptors (Supplementary Material S13).

For the case of k = 4, it was found that one cluster (No. 1) contained 79.6% of peptides with demonstrated Anti-SARS-CoV2 activity. When plotted using PCA , it was observed that these peptides were contained in a small cluster in the upper-left corner (Figure 4). This same cluster grouped a total of 1176 peptides out of 3155, representing only 37.3% of the total (Supplementary Material S14). Different from what happens with the grouping observed in the physicochemical space, with k=4, this rules out the possibility that the grouping results were due to chance (Table 5).

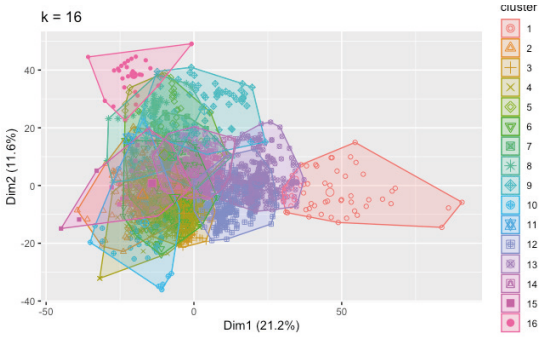


Fig. 3. PCA plot using the k-means algorithm at a level of k=16.

Cluster No. 14 contains 41.9% of the known Anti-SARS-CoV2.

k = 16	Total (%)	Anti SARS CoV2 (%)
Cluster 1	1,9	2,2
Cluster 2	7,4	4,3
Cluster 3	15,0	3,2
Cluster 4	6,0	5,4
Cluster 5	2,4	0,0
Cluster 6	8,6	9,7
Cluster 7	6,8	0,0
Cluster 8	2,7	1,1
Cluster 9	2,5	0,0
Cluster 10	1,8	0,0
Cluster 11	0,0	0,0
Cluster 12	9,4	21,5
Cluster 13	14,5	3,2
Cluster 14	**15,7**	**41,9**
Cluster 15	4,4	6,5
Cluster 16	1,0	1,1
	100	100

Table 3. Comparative results (Total Peptides vs Anti-SARS-CoV-2 Peptides) when using the k-means clustering algorithm with k=8.

Descriptors generated by StarPep			
Number of clusters k	**4**	**8**	**16**
Number of known Anti-SARS-CoV2 peptides in the cluster with the majority of known Anti-SARS-CoV2	47	55	39
Number of total peptides in the cluster with the majority of known Anti-SARS-CoV2	1556	687	495
% of known Anti-SARS-CoV2 peptides in the cluster with the majority of known Anti-SARS-CoV2	**50,5**	**59,1**	**41,9**
% of total peptides in the cluster with the majority of known Anti-SARS-CoV2	49,3	21,8	15,7
Do the Anti-SARS-Cov2 and Antiviral peptides show the same clustering pattern?	**YES**	**NO**	**NO**
Total of known Anti-SARS-CoV2 peptides = 93 Total peptides = 3155			

Table 4. Results of the peptide classification according to their Molecular Descriptors generated by StarPep, using the k-means algorithm with different k values.

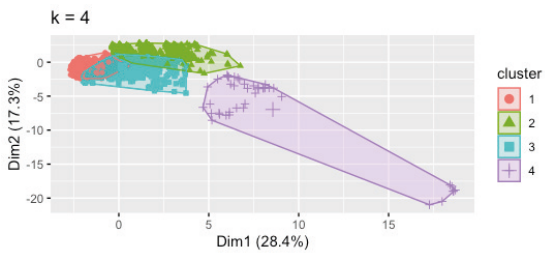


Fig. 4. Predictor's PCA plot using the k-means algorithm at a level of k=4. Cluster No. 1 contains 79.6% of the known Anti-SARS-CoV2.

k = 4	Total (%)	Anti-SARS-CoV2 (%)
Cluster 1	**37,3**	**79,6**
Cluster 2	27,4	12,9
Cluster 3	33,9	7,5
Cluster 4	1,4	0,0

Table 5. Predictor's results when using the k-means clustering algorithm with k=4.

In order to narrow down the number of potential peptides with Anti-SARS-CoV2 activity from this set, the value of k was increased to 8 and then to 16. In the case of k=8, cluster No. 4 grouped 51.6% of peptides with proven activity (Fig. 5). This same cluster contained a total of 236 peptides out of 3155 (Supplementary Material: S15), representing only 7.5% of the total (Table 6). It is inferred that these results cannot be attributed to chance, and it is demonstrated that by increasing the value of k from 4 to 8, the number of candidate peptides with potential anti-SARS-CoV2 activity is also reduced.

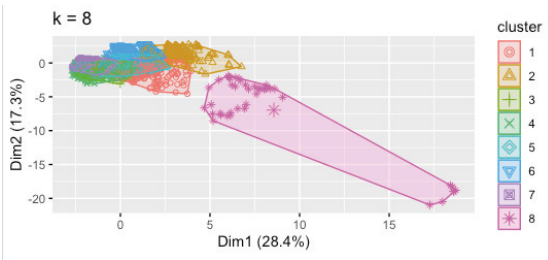


Fig. 5. Predictor's PCA plot using the k-means algorithm at a level of k=8. Cluster No. 4 contains 51.6% of the known Anti-SARS-CoV2.

k = 8	Total (%)	Anti-SARS-CoV2 (%)
Cluster 1	11,6	5,4
Cluster 2	12.9	4,3
Cluster 3	15,9	0.0
Cluster 4	7,5	**51,6**
Cluster 5	15,2	2,2
Cluster 6	13,6	8,6
Cluster 7	**22,0**	28,0
Cluster 8	1,4	0.0

Table 6. Predictor's results when using the k-means clustering algorithm with k=8

For k=16, it was observed that cluster No. 7 contained 50.5% of the known Anti-SARS-CoV2 peptides (Fig. 6), which corresponds to 47 peptides out of a total of 93 (Supplementary Material S16). Additionally, this cluster grouped 78 peptides (125 - 47) with potential Anti-SARS-CoV2 activity, representing only 4% of the total (Table 7). Similar to the previous cases, by using the predictions of AMP-Discover, these results do not appear to be due to chance. A summary of the results obtained

at different k levels for the clustering of peptides based on their predictions by AMPDiscover is presented in Table 8.

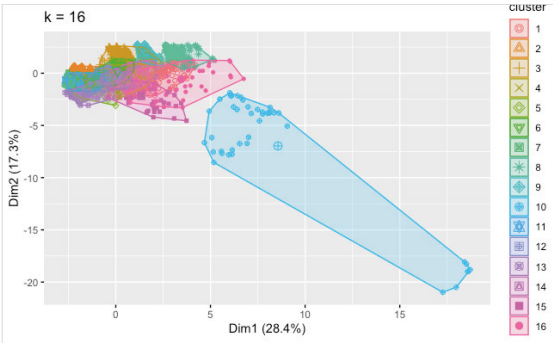


Fig. 6. Predictor's PCA plot using the k-means algorithm at a level of k=16 Cluster No. 7 contains 50.5% of the known Anti-SARS-CoV2.

k = 16	Total (%)	Anti SARS CoV2 (%)
Cluster 1	6,0	3,2
Cluster 2	13,6	14,0
Cluster 3	8,4	7,5
Cluster 4	5,5	0,0
Cluster 5	11,6	0,0
Cluster 6	6,0	0,0
Cluster 7	**4,0**	**50,5**
Cluster 8	8,0	3,2
Cluster 9	6,5	1,1
Cluster 10	1,4	0,0
Cluster 11	10,5	14,0
Cluster 12	5,4	0,0
Cluster 13	3,3	1,1
Cluster 14	5,8	1,1
Cluster 15	2,3	3,2
Cluster 16	1,9	1,1
	100	100

Table 7. Predictor's results when using the k-means clustering algorithm with k=16

Predictors generated by AMPDiscover			
Number of clusters k	**4**	**8**	**16**
Number of known Anti-SARS-CoV2 peptides in the cluster with the majority of known Anti-SARS-CoV2	74	48	47
Number of total peptides in the cluster with the majority of known Anti--SARS-CoV2	1176	236	125
% of known Anti-SARS-CoV2 peptides in the cluster with the majority of known Anti-SARS-CoV2	**79,6**	**51,6**	**50,5**
% of total peptides in the cluster with the majority of known Anti-SARS--CoV2	37,3	7,5	4,0
Has the clustering a random arrangement?	**NO**	**NO**	**NO**
Total of known Anti-SARS-CoV2 peptides = 93 Total peptides = 3155			

Table 8. Results of the peptide classification according to their predictions by AMPDiscover, using the k-means algorithm for different k values.

CLUSTERING ROBUSTNESS: PC DESCRIPTORS VS PREDICTORS

In order to study the robustness of the clustering when increasing k, a 3-level partition tree was constructed for k values of 4, 8, and 16. The results are shown in Figure 7 for Physicochemical (PC) descriptors (calculated by StarPep) and in Figure 8 for the predictor`s results (AMPDiscover). The sequence to cluster were obtained from the Anti-SARS-CoV-2 peptides clustering vector, at the mentioned levels of k, in both cases (Supplementary Material: S17 & Supplementary Material S18, respectively).

Figure 8 shows that at k=4, Cluster1 (C1) groups the majority of the known Anti-SARS--CoV-2 peptides (74 peptides out of 93); and, when it is further split into 2 clusters, at k=8, they contain all the elements of the first one. Similarly, when these 2 clusters are further divided at k=16, they show a subdivision containing all the elements of the previous level. On the contrary, in Figure 7, it is observed that the original cluster, containing the majority of the known Anti-SARS-CoV-2 (C2), was split in more than 2 clusters at k = 8. Additionally, at k = 16, all the clusters were split into more

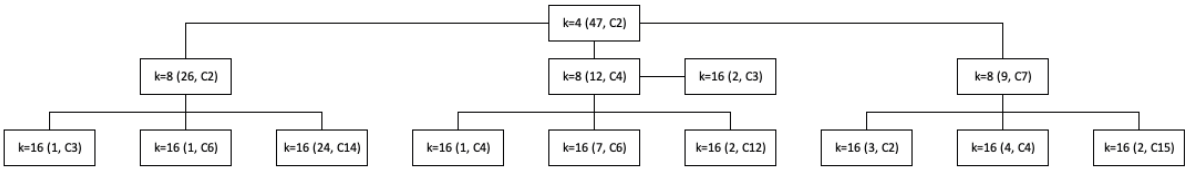


Figure 7. Physicochemical descriptors partition tree. Peptides clusters at different k values.

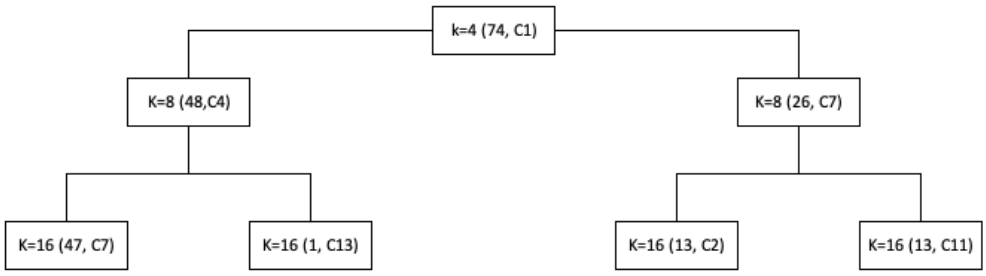


Figure 8. Predictors` metadescriptor partition tree. Peptides clusters at different k values. In the notation (X,Y), X is the number of sequences in cluster Y.

than 2 clusters. Moreover, the main cluster, when the sequences are represented by their physicochemical properties as computed by StarPep, contains only 47 elements, compared to 74 elements when the sequences are represented by their activity predictions in AMP-Discover.

Therefore, it is reasonable to assert that using features based on AMPDiscover predictions allows to build a better unsupervised learning algorithm compared to using physicochemical descriptors from StarPep. Specifically, to use them in combination with k-means, to identify and classify this kind of peptides. In addition, it is speculated that these predictions' partition results may be due to the fact that the peptides with known Anti--SARS-CoV-2 activity used as references have diverse mechanisms of action; therefore, the grouped peptides may also exhibit this diversity. However, this hypothesis needs to be demonstrated in further studies.

SEARCH FOR CANDIDATE PEPTIDES WITH ANTI-SARS-COV-2 ACTIVITY

In order to further reduce the number of candidate peptides with Anti-SARS-CoV-2 activity, the value of k was increased to 20. The set of descriptors generated by AMP-Discover was utilized for this clustering task (Supplementary Material S19). It was observed that peptides grouped in the resulting cluster, with the majority of the 93 known anti-SARS-CoV-2, were exactly the same as the peptides obtained using k=16. Moreover, the 47 Anti-SARS-CoV-2 grouped in this new cluster (C15) were exactly the same as the Anti-SARS-CoV-2 peptides grouped in C7 at k =16. So, it is not possible to continue reducing the potential candidate peptides through this filtering method. The summarized results of this run are presented in Table 9.

Predictors generated by AMPDiscover	
Level of k = **20**	AMPDiscover's Predictors
Number of known Anti-SARS-CoV2 peptides in the cluster with the majority of known Anti-SARS-CoV2	47
Number of total peptides in the cluster with the majority of known Anti-SARS-CoV2	125
% of known Anti-SARS-CoV2 peptides in the cluster with the majority of known Anti-SARS-CoV2	**50,5**
% of total peptides in the cluster with the majority of known Anti-SARS-CoV2	4,0
Has the clustering a random arrangement?	**NO**
Total of known Anti-SARS-CoV2 peptides = 93	
Total peptides = 3155	

Table 9. Results of the peptide classification according to their predictions, generated by AMPDiscover, using the k-means algorithm at k = 20

The study yielded, in total, 125 Anti-Virals in C15, the cluster with the majority of the Anti-SARS-CoV-2, at k=20 (Supplementary Material S20). Some of them (47) are well known Anti-SARS-CoV-2 peptides from the DRAMP database (Supplementary Material S21), but 78 peptides are new from StarPep's database (125 total Anti-Virals minus 47 known Anti-SARS-CoV-2), and correspond to less than 2.5% of the total StarPep's antivirals filtered (3155). So, it is plausible to say that these results are not due to chance, and these new 78 peptides could be candidates to have anti-SARS-CoV-2 activity. These candidate peptides are presented in Table 10.

The identified peptides were the input to ESM Fold[8] to predict their 3D structure and visualized using the Swiss-PdbViewer[9]. Some important structures of the tested Anti-SARS-CoV-2- peptide families are shown from Figure 10 to Figure 16. The rest of the EK1 & EKL1- peptide families' structures are presented in the Supplementary Material S22.

Analyzing the selected Anti-SARS-CoV-2 peptides in cluster 15, at k=20, it is remarkable that all of the peptides could be categorized. Some of them, belong to the EK1 & EKL1 families; others, were designed from the ACE--RBD interaction to block selectively of the Spike Protein (AHB1, Sap1, P8, and covid3 families); and the last, lipoproteins acting against the coronavirus Spike Protein. In addition, all of the peptide's structures are predicted to be a-helices, formed by 6 or more aminoacids, excepted by SAP2 and SAP6. However, SAP6 is the smallest's motif of a bigger a-helix structure (SAP1), and SAP2 has an overlapping region with SAP1 too (Larue et al., 2020).

Analyzing these peptides in more detail, it is found that EK1 can form a stable helix structure (Xia et al., 2019), inhibits SARS--CoV2 infection (Xia et al., 2020), and targets the HR1 domain in spike protein (Xia et al., 2020b). Recently, it has been found to be effective against the Omicron variant (Xia et al., 2022); especially, when it is linked to a palmitic acid to produce EK1-C16 (Lan et al., 2022). EKl1 is a lipopeptide, derived from EK1, which has been shown to be more potent and stable than the original (Zhou et al., 2022). Recently too, Lan et al., 2021, have designed EK1 peptides conjugated with 25-Hydrocholesterol showing inhibitory activity against SARS-CoV-2 and its variants, as well as other human coronaviruses.

AHB1 was designed against the RBD site (SARS-CoV2 Spike Protein), and blocks binding of ACE2 to RBD, but had low stability. AHB2, LCB1 & LCB3, an ACE2 derivatives too, were expressed and purified from *E. coli* and bind to RBD. AHB1 & AHB2 strongly neutralized infection of Vero E6 cells, but LCB1 & LCB3 showed a more potent neutralization (Cao et al., 2020).

Larue et al., 2020, rationally designed ACE-2-derivative peptides against SARS-CoV-2, too. They design six Spike-targeting ACE2-derived peptides (SAPs); SAP1, SAP2, SAP5 & SAP6 were effective. So, they demonstrate the feasibility of inhibiting this virus with peptide-based inhibitors.

Outlaw et al., 2020, designed a lipopeptide derived from the SARS-CoV-2 Spike glycoprotein HRC domain. This inhibits coronavirus cell entry in vitro & in vivo (Vero E6 monolayers), and has shown a broad spectrum. So, it could be a promising therapeutic agent against these viruses. This lipopeptide were more effective against an Ek1-derived peptide.

Valiente et al., 2021, computationally designed peptide inhibitors of SARS-CoV-2. They mimic the ACE2- a1- binding helix by searching a mirror-image version of the PDB. The two best designs bound the RBD and blocked the infection of Vero cells.

Another series of peptides acting mimicking the N-terminal helix of hACE2 protein, and containing the majority of the contacting residues, were design by Karoyan et al., 2021. These peptides (P8, P9 & P10) block SARS--CoV-2 pulmonary cells infection. They also showed a high helical folding in aqueous solution.

Finally, other lipoproteins resembling EK1, acting against the membrane fusion between the viral envelope and cell membrane, has been synthetized (Zhu et al., 2021). These includes MERS-LP & OC43-LP. They also design EKV1 & EKV2, other EK1-derivatives, but they have poor activity against divergent human CoVs.

Analyzing the selected 78 Anti-SARS--CoV-2 candidate peptides, it is possible to observe that 44 of them (56.41%) have a predicted *a*-helix in its structure. So, it may be worth to select them among the others for further studies (Figure 17). Is it possible, also, to continue discriminating peptides from this candidate list according to their ESM-Fold predicted structures. For example, some of the *a*-helix regions in the candidates are very small, less than 6 amino acids long; and, as the smallest reference peptide (Sap1) has more than this quantity, it is plausible to put them away from the candidate list for the time being. ESMFold could not predict a structure for the peptides which include an "X"

in its amino acid sequence (4/78 peptides), and they were discarded from the list. Table 11 shows the criteria & selection of the final 35 candidate peptides.

In order to understand better the selected candidate peptides and their antimicrobial activity, a search in DRAMP's database was performed. This search found that 5 of the 35 selected peptides has an antiviral activity documented in this database (Table 12). Moreover, three of them are synthetic constructions, and they have already some of their peptide sequences patented in the US as fusogenics (inhibiting fusion between cells and prevent viral replication). They are, also, peptides from human respiratory syncytial viruses, RSV. The other two, are peptides from natural sources; specifically, from juvenile cicadas (Cicadin; Wang & Ng, 2002) and Ginkgo biloba seeds (Ginkbilobin; Wang & Ng, 2000). They are antifungal peptides, but their Anti-SARS-CoV-2 activity has not been tested yet. So, for further research, this last two candidates could be highlighted to be tested against SARS-CoV2 in vitro. Finally, for their feasibility to be synthetized easily in the laboratory, the peptides with less than 20 aminoacids would be candidates to be tested in vitro, too. However, some previous *in silico* studies, using Molecular Docking & Molecular Dynamics, could be performed to save time and resources.

ID	Starpep_ID	Sequence	Biological Activity	PubMed_ ID
30	starPep_13393	VEGQLGENNELRLTRDAIE	AM, AV	16603508
40	starPep_13324	TVSTFIDLNITMLED	AM, AV	19104014
48	starPep_13097	SVERIKTTSSIEFAR	AM, AV	19104014
54	starPep_13076	STCVPVAADNVIVQN	AM, AV	19104014
55	starPep_13072	SSSTSTQVQILSNAL	AM, AV	8380075
105	starPep_12412	QRRNQLHDLRFADID	AM, AV	19104014
113	starPep_12188	PLVEGQLGENNELRL	AM, AV	19104014
116	starPep_12270	PSRVEAFHRYGTTVN	AM, AV	19104014
118	starPep_12227	PPPPGASANASVERI	AM, AV	19104014
121	starPep_12056	NNLETTAFHRDDHET	AM, AV	19104014
126	starPep_12010	NELRLTRDAIEPCTV	AM, AV	19104014
133	starPep_11326	LRSEYGGSFRFSSDA	AM, AV	19104014
136	starPep_11278	LQFTYNHIQRHVNDM	AM, AV	19104014
159	starPep_10268	IEFARLQFTYNHIQR	AM, AV	19104014
197	starPep_09732	GENNELRLTRDAI	AM, AV	16603508
204	starPep_09304	EVDEMLRSEYGGSFR	AM, AV	19104014
210	starPep_09169	DYTEVQRRNQLHDLR	AM, AV	19104014
215	starPep_09155	DVREEEQLGERATGLNLNI	AM, AV	16603508
216	starPep_09219	EEYAYSHQLSRADIT	AM, AV	19104014
227	starPep_09103	DMELKPANAATRTSR	AM, AV	19104014
400	starPep_12240	PPVYTKDVDISSQISSMNQSLQQSKDYIKEAQKILDTVNPSL	AM, AV	16973588
401	starPep_10982	KVDISSQISSMNQSLQQSKDYIKEAQRLLDTVNPSL	AM, AV	16026621
521	starPep_10422	IPESSELTLQELLGEERR	AM, AV	15182185
576	starPep_09164	DWVAVKQSYF	AM, AV	12917476
609	starPep_06548	LGQGVSIE	AM, AV, Anti-HIV	10074409
630	starPep_11998	NASDMEIKKVNKKIEEYIKKIEEVEKKLEEVNKK	AM, AV	21601229
631	starPep_10906	KQNAANILRLKESIAATNEAVHEV	AM, AV	12127571
641	starPep_13812	XLVLQTMX	AM, AV	
675	starPep_09533	FLDSKAELEKARKILSEVGRWY	AM, AV	8521809
678	starPep_13847	YADHTGLVRDNMAKLRERLKQRQQLFDSQQGWFEGWFN-RSPWFTT	AM, AV	21900723
694	starPep_13875	YDHIQDHVNTMFSRLATSWCLLQNKERALWAEAA	AM, AV	15269351
697	starPep_13638	WHSRGSTWLYRETANLNAMLTITTARSKYPY	AM, AV	21342525
747	starPep_13881	YEIIMDIEQNNVQGKTGIQQ	AM, AV	8661378
769	starPep_13544	VSVIVPDYQCYLDRVDTWLQ	AM, AV	8661378
770	starPep_13338	VAGLRQSLEQYQVVKQPDYL	AM, AV	8661378
772	starPep_13251	TMKVDDLIVHFNMTKAVEMV	AM, AV	8661378
801	starPep_11872	MPEVEGEEIQPQMELRRNGR	AM, AV	8661378
823	starPep_08528	ATHQEAIEKVTGALKINNLR	AM, AV	8661378
979	starPep_09269	ENKCNGTDAKVKLIKQELDKYKNAVTELQLLMQST	AM, AV	
993	starPep_10311	IKENKCNGTDAKVKLIKQELDKYKNAVTELQLLMQ	AM, AV	
1012	starPep_13968	YQDVNCTDVSTAIHADQLTP	AM, AV	
1037	starPep_12455	RDVSDFTDSVRDPKTSEILD	AM, AV	
1071	starPep_09098	DLSLDFEKLNVTLLDLTYEMNRIQDAIKKLNESYINLKE	AM, AV	15150417

1195	starPep_12378	QMRRKVELFTYMRFD	AM, AV	
1209	starPep_12236	PPSHSFRPESLERLHLLRRVLLLMRIVH	AM, AV	15220414
1242	starPep_13562	VTTAQETKRGRIQTKKEVSI	AM, AV	9311818
1356	starPep_13885	YEPLVRRRSELMGRRNPV	AM, AV, Anti-HIV	20718496
1427	starPep_07897	VRDQAEHLKT	AM, AV, Anti-HIV	16854053
1430	starPep_13577	VVTLTDTTNQKTELQAIYLA	AM, AV, Anti-HIV	15790559
1553	starPep_12999	SLKIPNLD	AM, AV, Anti-HIV	21850718
1643	starPep_12170	PKDGPSPGGTLMDLSERQEVSSVRSLSST	AM, AV, Anti-HIV	17919453
1662	starPep_12212	PPINNCMPLGTEVSEALGGA	AM, AV, Anti-HIV	21543477
1669	starPep_34003	NHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLEL-DKWASLWNWFNIKIL	Anti-HIV	
1671	starPep_12145	PDKWTVQPIVLPEKDSWTVN	AM, AV, Anti-HIV	15790559
1708	starPep_11401	MAGRSGDSDEELLKTVRLIKFLYQSNPPPS	AM, AV, Anti-HIV	17403681
1716	starPep_06689	LSSGRPDGFIHVQGHLQEVD	AM, AV, Anti-HIV	21543477
1797	starPep_10975	KTELQAIYLALQDSGLEVNI	AM, AV, Anti-HIV	15790559
1810	starPep_06205	KILEPFRKQNPDIVIYQYMD	AM, AV, Anti-HIV	15790559
1824	starPep_10584	KGSPAIFQSSMTKIL	AM, AV, Anti-HIV	15790559
1842	starPep_06119	ITFEDLLDYYGP	AM, AV, Anti-HIV	16041387, 18374356
1859	starPep_06103	IQAQPDQSESELVNQIIEQL	AM, AV, Anti-HIV	15790559
2094	starPep_09249	ELLEELKNEAVRHFP	AM, AV, Anti-HIV	17490682
2107	starPep_09102	DLYVGSDLEIGQHRTKIEEL	AM, AV, Anti-HIV	15790559
2111	starPep_09117	DQAEHLKTAVQMAVFIHNYKA	AM, AV, Anti-HIV	22742518
2142	starPep_15775	AVKTITLNLVSPSANRYATFLTEIRDNVRXRSLDYSHSGIDVI-GAPSSRDSXLNINFQSP	AV	1936243
2156	starPep_05029	APKEWMEWDREINNYTSLIHSLIKQGI	AM, AV, Anti-HIV	11118065
2194	starPep_04947	AGERIVDIIATDIQ	AM, AV, Anti-HIV	12643937
2289	starPep_11224	LLGDLLDDVTSIRHAVLQNRAAIDF	AM, AV	15564453
2331	starPep_13255	TNDCPNSSVVYEAADAIL	AM, AV	
2364	starPep_13015	SEDVVSSSMSYTFT	AM, AV	22452666
2368	starPep_12934	SEDVVSSSMSYTFT	AM, AV	14694985
2732	starPep_08322	AKDLEVVTSTYVLVEA	AM, AV	14694985
2749	starPep_13802	XGQHPAKSMDVRRIEGGEILLNQLAGRMIPKGTLTWSGKFP	AM, AV	
2819	starPep_13786	XARLSPTMVHPNGAQP	AM, AV	
2887	starPep_09293	ESGRIKKEEFAEIMKICSTIEELRRQK	AM, AV	23279951
2953	starPep_00827	NEYHGFVDKANNENKRKKQQGRDDFVVKPNNFANRRRK-DDYNENYYDDVDAADVV	AM, AB, AF, AV	11814612, 12126728
2985	starPep_01125	KQTENLADTY	AM, AF, AV, Anti--Cancer	15572193
3057	starPep_00580	ANTAFVSSAHNTQKIPAGAPFNRNLRAMLADLRQNAAFAG	AM, AB, AF, AV, Anti-HIV	

Table 10. Peptides with potential Anti-SARS-CoV-2 activity identified using the clustering results (k-means; k=20) from AMPDiscover Predictors.

AM = Anti-Microbial, AB = Anti-Bacterial, AF = Antifungal, AV= Anti-Viral.

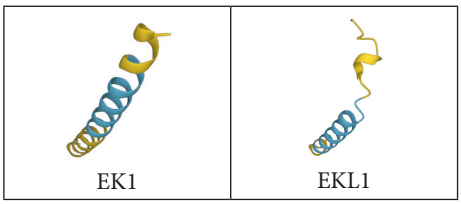


Figure 10. EK1 & EKl1 structures predicted by ESM Fold and visualized by Swiss-PdbViewer

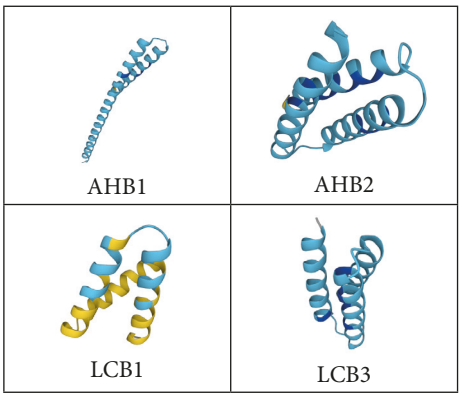


Figure 11. AHB1's Family structures predicted by ESM Fold and visualized by Swiss-PdbViewer

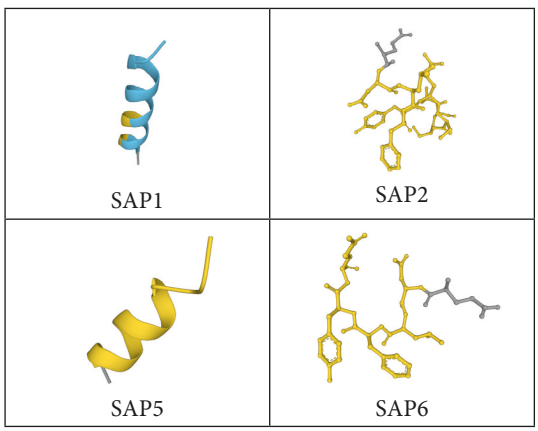


Figure 12. SAP's Family structures predicted by ESM Fold and visualized by Swiss-PdbViewer

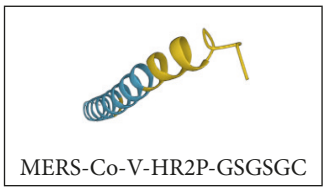


Figure 13. Lipoprotein structure, acting on the HRC Spike protein domain, predicted by ESM Fold and visualized by Swiss-PdbViewer

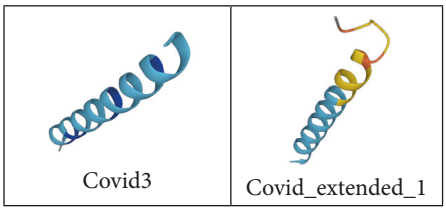


Figure 14. Covid's Family structures predicted by ESM Fold and visualized by Swiss-PdbViewer

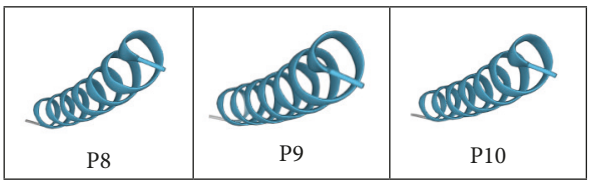


Figure 15. P-Peptides Family structures predicted by ESM Fold and visualized by Swiss-PdbViewer

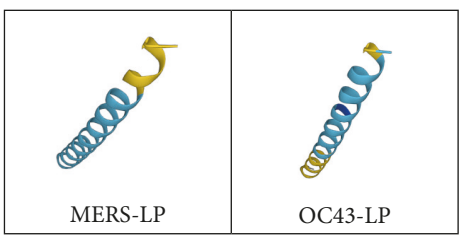
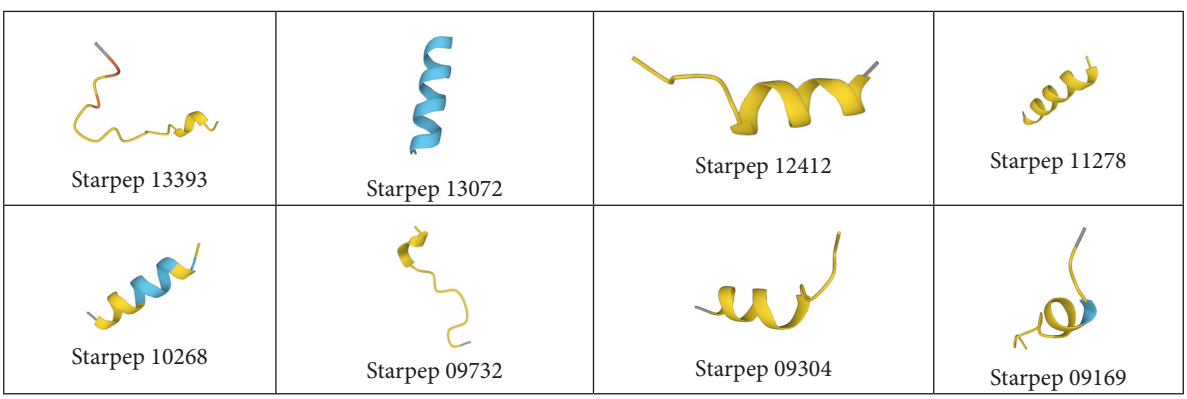


Figure 16. Anti-SARS-CoV2 Lipoproteins structure predicted by ESM Fold and visualized by Swiss-PdbViewer



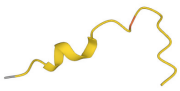
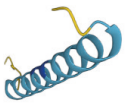
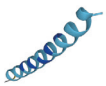
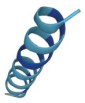
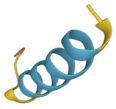
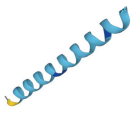
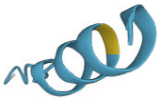
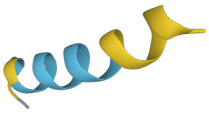
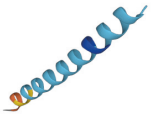
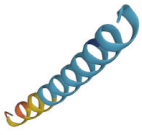
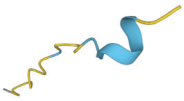
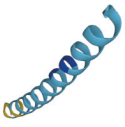
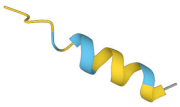
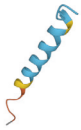
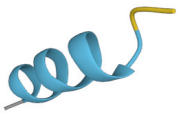
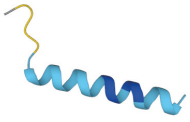
 Starpep 09155	 Starpep 09219	 Starpep 12240	 Starpep 10982
 Starpep 10422	 Starpep 11998	 Starpep 10906	 Starpep 09533
 Starpep 13847	 Starpep 13875	 Starpep 13638	 Starpep 13554
 Starpep 13338	 Starpep 13251	 Starpep 08528	 Starpep 09269
 Starpep 10311	 Starpep 13968	 Starpep 09098	 Starpep 12236
 Starpep 13885	 Starpep 13577	 Starpep 12170	 Starpep 12212
 Starpep 34003	 Starpep 11401	 Starpep 10975	 Starpep 06119
 Starpep 06103	 Starpep 09249	 Starpep 09117	 Starpep 05029
 Starpep 11224	 Starpep 29293	 Starpep 00827	 Starpep 00580

Figure 17. a-Helix's Candidate peptides, with potential Anti-SARS-CoV-2 activity, resulting from the use of the k-means clustering algorithm at k=20, based on AMPDiscover Predictors.

Starpep_ID	a-helix	a-helix, with 6 or more aa	Selected as Candidates
starPep_13393	Yes	No	
starPep_13324	No	No	
starPep_13097	No	No	
starPep_13076	No	No	
starPep_13072	Yes	Yes	Yes
starPep_12412	Yes	Yes	Yes
starPep_12188	No	No	
starPep_12270	No	No	
starPep_12227	No	No	
starPep_12056	No	No	
starPep_12010	No	No	
starPep_11326	No	No	
starPep_11278	Yes	Yes	Yes
starPep_10268	Yes	Yes	Yes
starPep_09732	Yes	No	
starPep_09304	Yes	Yes	Yes
starPep_09169	Yes	No	
starPep_09155	Yes	No	
starPep_09219	Yes	No	
starPep_09103	No	No	
starPep_12240	Yes	Yes	Yes
starPep_10982	Yes	Yes	Yes
starPep_10422	Yes	No	
starPep_09164	No	No	
starPep_06548	No	No	
starPep_11998	Yes	Yes	Yes
starPep_10906	Yes	Yes	Yes
starPep_13812	-	-	
starPep_09533	Yes	Yes	Yes
starPep_13847	Yes	Yes	Yes
starPep_13875	Yes	Yes	Yes
starPep_13638	Yes	Yes	Yes
starPep_13881	No	No	
starPep_13544	Yes	Yes	Yes
starPep_13338	Yes	Yes	Yes
starPep_13251	Yes	Yes	Yes
starPep_11872	No	No	
starPep_08528	Yes	Yes	Yes
starPep_09269	Yes	Yes	Yes
starPep_10311	Yes	Yes	Yes
starPep_13968	Yes	No	
starPep_12455	No	No	
starPep_09098	Yes	Yes	Yes
starPep_12378	No	No	
starPep_12236	Yes	Yes	Yes
starPep_13562	No	No	
starPep_13885	Yes	Yes	Yes
starPep_07897	No	No	
starPep_13577	Yes	Yes	Yes
starPep_12999	No	No	
starPep_12170	Yes	No	
starPep_12212	Yes	No	
starPep_34003	Yes	Yes	Yes
starPep_12145	No	No	
starPep_11401	Yes	Yes	Yes
starPep_06689	No	No	
starPep_10975	Yes	Yes	Yes
starPep_06205	No	No	
starPep_10584	No	No	
starPep_06119	Yes	Yes	Yes
starPep_06103	Yes	Yes	Yes
starPep_09249	Yes	Yes	Yes
starPep_09102	No	No	
starPep_09117	Yes	Yes	Yes
starPep_15775	-	-	
starPep_05029	Yes	Yes	Yes
starPep_04947	No	No	
starPep_11224	Yes	Yes	Yes
starPep_13255	No	No	
starPep_13015	No	No	
starPep_12934	No	No	
starPep_08322	No	No	
starPep_13802	-	-	
starPep_13786	-	-	
starPep_09293	Yes	Yes	Yes
starPep_00827	Yes	Yes	Yes
starPep_01125	No	No	No
starPep_00580	Yes	Yes	Yes

Table 11. Selection of the 35 final Candidate Peptides for further studies, according to a subjective criterion based in EMS Fold predictions

Starpep_ID	DRAMP ID	Peptide Name	Source	Activity	PMID/Patent Number
starPep_09269	DRAMP16181	Sequence from Patent US 6228983	Synthetic construct	AM, AV	US 6228983 B1
starPep_10311	DRAMP16181	Sequence from Patent US 6228983	Synthetic construct	AM, AV	US 6228983 B1
starPep_06119	DRAMP15983	Sequence 1 from Patent US 20100130430	Synthetic construct	AM, AV	US 2010/0130430 A1
starPep_00827	DRAMP03465	Cicadin (Insects, animals)		AM, AF, AV	11814612
starPep_00580	DRAMP00341	Antifungal protein ginkbilobin-1 (Ginkbilobin, GNL; Plants)	Ginkgo biloba (Ginkgo) (Maidenhair tree)	AM, AB, AF, AV	

Table 12. Candidate Peptides founded in DRAMP's database with Antiviral activity. AM = Anti-Microbial, AB = Anti-Bacterial, AF = Antifungal, AV= Anti-Viral.

CONCLUSIONS

In this work, Anti-SARS-CoV-2 peptides were gathered from the literature and the DRAMP database, and their molecular descriptors were computed by using the StarPep Toolbox. The same set of selected peptides were fed to the AMP Discover predictor. We used both representations and compared them in their ability to cluster similar peptides with respect to their activities by using the k-means algorithm. The results obtained through k-means clustering and Principal Component Analysis (PCA) techniques indicate an apparent grouping pattern of the tested Anti-SARS-CoV2 peptides in both cases. This suggests that, it is possible to identify peptides with potential Anti-SARS-CoV-2 activity through specific clusters.

In addition, the representation generated by the predictor shows that when increasing the number of clusters k, from 4 to 8, and from 8 to 16, all the elements contained in the mother cluster goes to one of two offspring. From these results it is conjectured that the grouping pattern of Anti-SARS-CoV-2 peptides is biased by their structure. As the majority of the witness peptides in the cluster with majority of Anti-SARS-Cov2 belongs to the EKL1 family, it is inferred that targeting the Spike glycoprotein could be the mechanism of action of the candidate peptides.

This work has contributed to show a plausible methodology to follow to find AVPs with possible Anti-SARS-CoV-2 activity, using predictions obtained from AMPDiscover, in combination with a non-supervised algorithm (k-means). It has been shown, also, that the results from the predictors (AMPDiscover) are more robust than the obtained with physical-chemical descriptors (StarPep), when the number of clusters k is increased. Finally, it was observed that some of the Anti-SARS-CoV2 peptides could be grouped according to their structure and their ability to form α-helices. Hence, is it possible that the selected candidate peptides in the same group, obtained from the methodology presented here, share a mechanism of action against the coronavirus Spike protein, as some of the Anti-SARS-CoV2 reference peptides do.

REFERENCES

Aguilera-Mendoza, L., Ayala-Ruano, S., Martinez-Rios, F., Chavez, E., García-Jacas, C. R., Brizuela, C. A., & Marrero-Ponce, Y. (2023). StarPep Toolbox: an open-source software to assist chemical space analysis of bioactive peptides and their functions using complex networks. Bioinformatics, 39(8), btad506.

Caly, L., Druce, J. D., Catton, M. G., Jans, D. A., & Wagstaff, K. M. (2020). The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. Antiviral research, 178, 104787.

Cao, L., Goreshnik, I., Coventry, B., Case, J. B., Miller, L., Kozodoy, L., ... & Baker, D. (2020). De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science, 370(6515), 426-431.

Diamond, G., Beckloff, N., Weinberg, A., & Kisich, K. O. (2009). The roles of antimicrobial peptides in innate host defense. Current pharmaceutical design, 15(21), 2377-2392.

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV—a target for vaccine and therapeutic development. Nature Reviews Microbiology, 7(3), 226-236.

Essa, R. Z., Wu, Y. S., Batumalaie, K., Sekar, M., & Poh, C. L. (2022). Antiviral peptides against SARS-CoV-2: Therapeutic targets, mechanistic antiviral activity, and efficient delivery. Pharmacological Reports, 74(6), 1166-1181.

Fakih, T. M., Dewi, M. L., & Syahroni, E. (2022). The Inhibition of Angiotensin-Converting Enzyme 2 Receptors of SARS-CoV-2 Through Mucroporin Derived from Scorpion Venom. KnE Life Sciences, 92-102.

Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., ... & Flanigan, T. (2020). Compassionate use of remdesivir for patients with severe Covid-19. *New England Journal of Medicine*, *382*(24), 2327-2336.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. cell, 181(2), 271-280.

Karoyan, P., Vieillard, V., Gómez-Morales, L., Odile, E., Guihot, A., Luyt, C. E., ... & Lequin, O. (2021). Human ACE2 peptide-mimics block SARS-CoV-2 pulmonary cells infection. *Communications biology*, *4*(1), 197.

Lan, Q., Chan, J. F. W., Xu, W., Wang, L., Jiao, F., Zhang, G., ... & Wang, Q. (2022). A palmitic acid-conjugated, peptide-based pan-CoV fusion inhibitor potently inhibits infection of SARS-CoV-2 Omicron and other variants of concern. *Viruses*, *14*(3), 549.

Lan, Q., Wang, C., Zhou, J., Wang, L., Jiao, F., Zhang, Y., ... & Jiang, S. (2021). 25-Hydroxycholesterol-conjugated EK1 peptide with potent and broad-spectrum inhibitory activity against SARS-CoV-2, its variants of concern, and other human Coronaviruses. *International journal of molecular sciences*, *22*(21), 11869.

Larue, R. C., Xing, E., Kenney, A. D., Zhang, Y., Tuazon, J. A., Li, J., ... & Sharma, A. (2020). Rationally designed ACE2-derived peptides inhibit SARS-CoV-2. *Bioconjugate Chemistry*, *32*(1), 215-223.

Li, Q., Zhao, Z., Zhou, D., Chen, Y., Hong, W., Cao, L., ... & Li, W. (2011). Virucidal activity of a scorpion venom peptide variant mucroporin-M1 against measles, SARS-CoV and influenza H5N1 viruses. Peptides, 32(7), 1518-1525.

Ledford H. (2020). Coronavirus breakthrough: dexamethasone is first drug shown to save lives. *Nature*, *582*(7813), 469.

Ling, R., Dai, Y., Huang, B., Huang, W., Yu, J., Lu, X., & Jiang, Y. (2020). In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2. Peptides, 130, 170328.

Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C. R., ... & Jiang, S. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. The Lancet, 363(9413), 938-947.

Mahendran, A. S. K., Lim, Y. S., Fang, C. M., Loh, H. S., & Le, C. F. (2020). The potential of antiviral peptides as COVID-19 therapeutics. Frontiers in pharmacology, 11, 575444.

Niv, Y. (2020). Defensin 5 for prevention of SARS-CoV-2 invasion and Covid-19 disease. Medical hypotheses, 143, 110244.

Ohashi, H., Watashi, K., Saso, W., Shionoya, K., Iwanami, S., Hirokawa, T., ... & Wakita, T. (2020). Multidrug treatment with nelfinavir and cepharanthine against COVID-19. *BioRxiv*, 2020-04.

Outlaw, V. K., Bovier, F. T., Mears, M. C., Cajimat, M. N., Zhu, Y., Lin, M. J., ... & Porotto, M. (2020). Inhibition of coronavirus entry in vitro and ex vivo by a lipid-conjugated peptide derived from the SARS-CoV-2 spike glycoprotein HRC domain. *MBio*, *11*(5), 10-1128.

Pinacho-Castellanos, S. A., García-Jacas, C. R., Gilson, M. K., & Brizuela, C. A. (2021). Alignment-free antimicrobial peptide predictors: improving performance by a thorough analysis of the largest available data set. Journal of Chemical Information and Modeling, 61(6), 3141-3157.

Rani, P., Kapoor, B., Gulati, M., Atanasov, A. G., Alzahrani, Q., & Gupta, R. (2022). Antimicrobial peptides: A plausible approach for COVID-19 treatment. Expert opinion on drug discovery, 17(5), 473-487.

Romero-Molina, S., Ruiz-Blanco, Y. B., Green, J. R., & Sanchez-Garcia, E. (2019). ProtDCal-Suite: a web server for the numerical codification and functional analysis of proteins. *Protein Science*, *28*(9), 1734-1743.

Souza, F. R., Moura, P. G., Costa, R. K. M., Silva, R. S., & Pimentel, A. S. (2023). Absolute binding free energies of mucroporin and its analog mucroporin-M1 with the heptad repeat 1 domain and RNA-dependent RNA polymerase of SARS-CoV-2. Journal of Biomolecular Structure and Dynamics, 41(14), 6957-6968.

Valiente, P. A., Wen, H., Nim, S., Lee, J., Kim, H. J., Kim, J., ... & Kim, P. M. (2021). Computational design of potent D-peptide inhibitors of SARS-CoV-2. *Journal of Medicinal Chemistry*, *64*(20), 14955-14967.

Wang, K., Chen, W., Zhou, Y. S., Lian, J. Q., Zhang, Z., Du, P., ... & Chen, Z. N. (2020). SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. biorxiv, 2020-03.

Wang, J., Dou, X., Song, J., Lyu, Y., Zhu, X., Xu, L., ... & Shan, A. (2019). Antimicrobial peptides: Promising alternatives in the post feeding antibiotic era. Medicinal research reviews, 39(3), 831-859.

Wang, H., & Ng, T. B. (2002). Isolation of cicadin, a novel and potent antifungal peptide from dried juvenile cicadas. *Peptides*, *23*(1), 7-11.

Wang, H., & Ng, T. B. (2000). Ginkbilobin, a novel antifungal protein from Ginkgo biloba seeds with sequence similarity to embryo-abundant protein. *Biochemical and Biophysical Research Communications*, *279*(2), 407-411.

Wang, H., Yang, P., Liu, K., Guo, F., Zhang, Y., Zhang, G., & Jiang, C. (2008). SARS coronavirus entry into host cells through a novel clathrin-and caveolae-independent endocytic pathway. *Cell research*, *18*(2), 290-301.

Wohlford-Lenane, C. L., Meyerholz, D. K., Perlman, S., Zhou, H., Tran, D., Selsted, M. E., & McCray Jr, P. B. (2009). Rhesus theta-defensin prevents death in a mouse model of severe acute respiratory syndrome coronavirus pulmonary disease. Journal of virology, 83(21), 11385-11390.

World Health Organization. (2020). COVID-19 weekly epidemiological update, 3 November 2020.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science, 367(6483), 1260-1263.

Xia, S., Chan, J. F. W., Wang, L., Jiao, F., Chik, K. K. H., Chu, H., ... & Jiang, S. (2022). Peptide-based pan-CoV fusion inhibitors maintain high potency against SARS-CoV-2 Omicron variant. *Cell Research*, *32*(4), 404-406.

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., ... & Lu, L. (2020a). Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell research*, *30*(4), 343-355.

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., ... & Lu, L. (2020b). Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cellular & molecular immunology, 17(7), 765-767.

Xia, S., Yan, L., Xu, W., Agrawal, A. S., Algaissi, A., Tseng, C. T. K., ... & Lu, L. (2019). A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. Science advances, 5(4), eaav4580.

Zhang, Q. Y., Yan, Z. B., Meng, Y. M., Hong, X. Y., Shao, G., Ma, J. J., ... & Fu, C. Y. (2021). Antimicrobial peptides: mechanism of action, activity and clinical potential. Military Medical Research, 8, 1-25.

Zhang, L. J., & Gallo, R. L. (2016). Antimicrobial peptides. Current Biology, 26(1), R14-R19.

Zhao, H., To, K. K., Lam, H., Zhou, X., Chan, J. F. W., Peng, Z., ... & Yuen, K. Y. (2021). Cross-linking peptide and repurposed drugs inhibit both entry pathways of SARS-CoV-2. Nature communications, 12(1), 1517

Zhao, H., To, K. K., Sze, K. H., Yung, T. T. M., Bian, M., Lam, H., ... & Yuen, K. Y. (2020). A broad-spectrum virus-and host-targeting peptide against respiratory viruses including influenza virus and SARS-CoV-2. Nature communications, 11(1), 4252.

Zhao, H., Zhou, J., Zhang, K., Chu, H., Liu, D., Poon, V. K. M., ... & Zheng, B. J. (2016). A novel peptide with potent and broad-spectrum antiviral activities against multiple respiratory viruses. Scientific reports, 6(1), 22008.

Zhou, J., Xu, W., Liu, Z., Wang, C., Xia, S., Lan, Q., ... & Wang, Q. (2022). A highly potent and stable pan-coronavirus fusion inhibitor as a candidate prophylactic and therapeutic for COVID-19 and other coronavirus diseases. *Acta Pharmaceutica Sinica B*, *12*(4), 1652-1661.

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. nature, 579(7798), 270-273.

Zhu, Y., Yu, D., Hu, Y., Wu, T., Chong, H., & He, Y. (2021). SARS-CoV-2-derived fusion inhibitor lipopeptides exhibit highly potent and broad-spectrum activity against divergent human coronaviruses. *Signal transduction and targeted therapy*, *6*(1), 294.

WEB PAGES:

1. https://www.worldometers.info/coronavirus/

2. http://dramp.cpu-bioinfor.org/

3. http://mobiosd-hub.com/starpep

4. https://biocom-ampdiscover.cicese.mx

5. https://cran.rstudio.com/

6. https://posit.co/download/rstudio-desktop/

7. https://rpubs.com/

8. https://esmatlas.com/resources?action=fold

9. https://spdbv.unil.ch/disclaim.html#