Journal of AIDS & Clinical Research
ISSN: 2155-6113

Tincho et al., J AIDS Clin Res 2016, 7:9
DOI: 10.4172/2155-6113.1000606

Research Article Open Access

In Silico Identification and Molecular Validation of Putative Antimicrobial Peptides for HIV Therapy

Marius Belmondo Tincho[1], Musa Nur Gabere[2] and Ashley Pretorius[1*]

[1]*Department of Biotechnology, University of the Western Cape, South Africa*
[2]*Department of Bioinformatics, King Abdullah International Medical Research Center, Saudi Arabia*

Abstract

Objective: Despite the effort to reduce the rate of HIV infection, AIDS still remains the major cause of death around the world, predominantly in Sub-Sahara Africa. Neither a cure, nor an HIV vaccine has been found to date and the disease can only be managed by using High Active Antiretroviral Therapy (HAART). The need for non-toxic regiments has brought about the necessity for additional HIV treatment to lower mortality rates. Antimicrobial Peptides (AMPs) had proven to be a promising therapeutic agent against HIV. The aim of this research was to identify AMPs, which binds gp120 at the area where gp120 interacts with CD4+, to prevent HIV invasion and HIV replication.

Method: Putative AMPs were identified using an *In Silico* mathematical algorithm, Profile Hidden Markov Models (HMMER). The AMPS 3-D structures was carried out using I-TASSER and the modelled AMPs were docked against the HIV protein gp120 using PATCHDOCK. Subsequently, molecular method was used to show the anti-HIV ability of these putative to validate by inhibiting HIV-1 replication.

Results: The *In Silico* results showed that 30 putative anti-HIV AMPs were identified. Furthermore, out of the 10 best ranked putative AMPs, based on their E-value, selected for *In Silico* docking, two AMPs proved to inhibit HIV-1 NL4-3 with maximal effective concentration (EC_{50}) values of 37.5 µg/ml and 93.75 µg/ml respectively. This result looks promising since 150 µg/ml AMPs could not achieved 80% toxicity of the human T cells, thus high Therapeutics Index (TI) might be obtained if 50% cytotoxic concentration (CC_{50}) is established.

Conclusion: The ability of these AMPs to inhibit HIV replication justifies the usage of HMMER in design and discovery. Additionally, these AMPs pave the way for the design of anti-HIV peptide-based drugs.

Keywords: Human immunodeficiency virus; HIV pandemic; HIV treatment regimens; gp120 protein; T cells; CD4; Anti-HIV antimicrobial peptides; *In Silico* approach; Profile hidden Markov models (HMMER); *In Silico* protein-peptide interaction

Introduction

Since the discovery of Human Immunodeficiency Virus (HIV) more than 30 years ago, the Acquired Immunodeficiency Syndrome (AIDS) is considered as one of the major clinical diseases and a health problem around the globe, especially in Sub-Saharan Africa [1]. The driving force behind the development of High Active Antiretroviral Therapy (HAART) for HIV treatment has been inevitable due to the emergence of HIV progression. This demand has continuously been met by the pharmaceutical industry by developing new drugs or by modifying the existing medication in a timely fashion. Notwithstanding these advancements, the rapid emergence of resistance to current treatment regimens is even a greater problem for life-threatening HIV infections due to single class of HAART treatment and non-adherence to medication, therefore a daunting clinical problem. The only effective solution to this problem would thus seem to develop a combination of therapies involving several anti-HIV regimens with different mechanisms of inhibitory action.

Although AMPs generally exhibit lower potency against susceptible microbial targets compared to conventional low molecular weight antibiotic compounds, they hold several compensatory advantages including: (*i*) fast killing (*ii*) broad range of activity (*iii*) low toxicity and (*iv*) minimal development of resistance in target organisms [2]. Thus, AMPs have drawn significant attention as possible sources of novel antimicrobial agents specifically against HIV/AIDS [3]. Several studies have investigated the anti-HIV activity of Human Neutrophil Peptides (HNP1, HNP2, HNP3 and HNP4) and all showed activities against HIV primary isolates, using two mechanisms. Firstly, they can inhibit HIV-1 replication by a direct interaction with the virus as well as by affecting the target cells [4,5]. Chang et al. showed that in the absence of serum, HNP1 could directly inactivate the virus before it infects a cell. Whilst in infected cells and in the presence of serum and non-cytotoxic concentrations (low dose), HNP1 blocks HIV-1 infection at the steps of nuclear import and transcription. Secondly, in primary CD4+ T cells, HNP1 interference with PKC signaling is associated with the ability of HNP1 to inhibit infection after HIV enters the cell [4]. In addition, another of the member of the group, HNP4, acts in a lectin-independent manner thus has no binding affinity either to the HIV envelope glycoprotein gp120 or CD4+ [5]. However, HNP4 inhibits HIV replication more effectively [6]. Furthermore, other potent anti-HIV AMPs have been reported in great detail [7-9].

It stands to reason as the demand increases for the identification of AMPs, parallel technologies are developed to meet this demand. With the birth of Bioinformatics, a host of technologies using *In Silico* approaches to identify AMPs has been established and have promised

***Corresponding author:** Ashley Pretorius (PhD), Department of Biotechnology, University of the Western Cape, Life Science Building (Room 2.48), Bellville, South Africa, Tel: + 27 (0) 21 959 9727; +27 (0) 82 329 5646; Fax: + 27 (0) 21 959 3505; E-mail: aspretorius@uwc.ac.za

Received July 05, 2016; **Accepted** July 30, 2016; **Published** August 06, 2016

Citation: Tincho MB, Gabere MN, Pretorius A (2016) *In Silico* Identification and Molecular Validation of Putative Antimicrobial Peptides for HIV Therapy. J AIDS Clin Res 7: 606. doi:10.4172/2155-6113.1000606

to be less time consuming, more cost effective and less labour intensive thus speeding up the discovery process. The creation of various AMP databases such as Bactibase [10,11], APD2 [12,13], AMPer [14], CAMP [15] and DAMPD [16] has further boosted these prediction methods. Besides hosting a wealth of AMPs, these databases also incorporate many embedded algorithms for AMP identification thus providing an indispensable knowledge-based for both qualitative and quantitative activity prediction models using tools such as Support Vector Machine (SVM) [15], profile Hidden Markov Models (HMMER) [14,17], Gap Local Alignment of Motifs 2 (GLAM2) [18], Quantitative Structure Activity Relationship (QSAR) [19], Linear Discriminant Analysis (LD), Random Forest (RF) [15] and Sliding Window (SW) [20]. Such approach has allowed for the systematic mining of genomic expressed sequence tag data, the aim of which was to discover hitherto un-described natural AMP sequence [21].

Prediction models can also be constructed independently, with the incorporation of properties related to the structure and the activity of existing peptides using the same tools. Whilst SVM, QSAR, LD and RF require structure-activity relationship information in order to enhance their strength and performances [20], HMMER and GLAM2 however only require the sequences of experimentally validated biomolecules for the construction of models. These models will display their features based on the motifs of the input sequences and would have the desired activity against the specific target [18,22]. Here, we report the construction of HMMER profiles using experimentally validated anti-HIV AMPs from various AMPs databases. Furthermore, we tested the robustness of each HMMER profiles by scanning it against an anti-HIV AMP testing set and evaluate its performance, and the algorithm ability's to identify and highly discriminate putative AMPs from non-AMPs. *In Silico* protein-peptide interaction was performed to screen for possible putative AMPs that will have gp120 as receptor, after predicting the 3-Dimensional structure of these novel AMPs. Finally, anti-HIV activity of the putative AMPs was screened for since it was hypothesized from the docking study that these AMPs interact with the gp120 protein at the site of gp120's interaction with CD4+ and could serve as an inhibitory molecule for the gp120-CD4+ interaction. Thus, these putative AMPs could inhibit HIV replication and would be utilised as potent lead compounds to formulate entry inhibitor peptide-based drugs.

Materials and Methods

Data retrieval

The experimentally validated anti-HIV AMPs were retrieved from antimicrobial peptide databases namely: Antimicrobial Peptides Database (APD) [12,13], Collection of Antimicrobial Peptides (CAMP) [15], Cybase [23,24], UniprotKB [25] and Dragon Anti-Microbial Peptide Database (DAMPD) [16]. Thereafter, curation was performed to verify that all the retrieved anti-HIV AMPs were either experimentally validated or predicted. Duplicate experimentally validated anti-HIV AMPs were then removed from the list using Cluster Database at High Identity with Tolerance (CD-HIT) [26].

Training and testing data sets

The final list of the experimentally validated anti-HIV AMPs was classified according to their super-family: Amphibian, Microorganism, Human defensin, Fish and Crab, Insect, Vertebrate and Plant super-families. The need to divide these peptides into their various families is as a result of their diversity, in terms of sequence and activity [23,27,28]. Each super-family AMP data set was then divided into two portions: three-quarters of each super-family data set was utilised as the training set whilst one-quarter was used as the testing set.

Software for construction AMPs profiles

The Hidden Markov Models (HMMER) algorithm version 2.3.2 [21] was used to construct seven super-family (Amphibians, Microorganisms, Human, Defensins, Fish and Crabs, Insects, Vertebrates and Plants) models/profiles using the respective training sets. All the HMMER profiles were constructed using Ubuntu 12.04 LTS operating system, which is based on the Linux kernel. To construct these profiles, the training data set from the experimentally validated anti-HIV AMPs was used to construct the profiles or in the training of the profiles. Furthermore, the testing data set was used for validating the robustness of each profile created. The task was accomplished on a terminal and the command lines used to build each profile was written in accordance with the corresponding algorithm and the steps involved in their construction were as below:

The profile HMMER has multiple modules for it to perform optimally. For the first step, the training sets of each super-family were aligned using the ClustalW alignment tool [29]. The alignment was performed using the command line:

clustalw–align–output=gcg–case=upper–sequos=off–outorder=aligned–infile=family.fasta ….

The command line simply stated <<do an alignment of the sequences which are in the upper case found in the input file "family.fasta" with the FastA, using ClustalW as multiple alignment tool and GCG Postscript output for graphical printing>>. The output of the command results in the creation of aligned sequences, called "family.msf". This aligned sequence was used as input in the next step.

The next step enables the creation of profiles/signatures of family sequences by showing the common motifs within the model. To achieve this, the "Build profiles" was run using the following command:

hmmbuild family.hmm family.msf…… ……

To enhance the sensitivity of the profile, the file generated (family.hmm) from the profile building was calibrated by using the command line:

hmmcalibrate family.hmm ……..

The generated profiles "family.hmm" was used in assessing their performance by testing an independent AMP dataset. In total, seven family AMP profiles were created namely, Amphibians, Microorganisms, Human defensins, Fish and crabs, Insects, Vertebrates and Plants.

Independent model testing

The independent testing of each profile was performed for each family in a step called "Query profiles". The query of profiles (in FASTA format) also confirmed that the testing and the training sets had anti-HIV activity since both sets were derived from the list of experimentally validated anti-HIV AMPs. Independent testing was done with the testing set of each profile specific to their super-family, which represented about a quarter of the data set. The testing data were queried against the created profiles using the command line, with an E-value threshold of 5% or 0.05:

hmmsearch–E 5e-2 family.hmm familyquery.txt>resultfile.txt …… ..

Performance measures of each profile based on prediction of the positive and the negative testing set

Statistical performance measures were then calculated using sensitivity, specificity, accuracy and Mathew Correlation Coefficient for each profile. In addition to the positive testing mentioned in the

previous section, a negative data set consisting of 596 neuro-peptide sequences, which are non anti-HIV AMPs, was also used to measure the performance of each profile. Note that we need information on four statistics that is *TP, TN, FP* and *FN*. *TP* (True positive) represents correctly predicted positive examples (anti-HIV AMPs), *TN* (True negative) is correctly predicted negative examples (non-anti-HIV AMPs), *FP* (False positive) is the number of non-anti-HIV AMPs examples wrongly predicted as anti-HIV AMPs, *FN* (False negative) is the number of anti-HIV AMPs wrongly predicted as non-anti-HIV AMPs. The measures used are described as follows:

- **Sensitivity** is the percentage of anti-HIV AMPs (testing sets) correctly predicted as anti-HIV AMPs (positive). The sensitivity (recall) is defined in equation (1):

$$Sensitivity = \left(\frac{TP}{TP+FN}\right)\times 100 \quad (1)$$

- **Specificity** is the percentage of non-anti-HIV AMPs (negative sets) correctly predicted as non-anti-HIV AMPs (negative). The specificity is defined in equation (2):

$$Specificity = \left(\frac{TN}{TN+FP}\right)\times 100 \quad (2)$$

- **Accuracy** is the percentage of correctly predicted peptides (anti-HIV AMPs and non-anti-HIV AMPs). The accuracy is defined in equation (3):

$$Accuracy = \left(\frac{TP+TN}{TP+FP+TN+FN}\right)\times 100 \quad (3)$$

- **Mathew's correlation coefficient (MCC)** is a measure of both sensitivity and specificity. MCC=0 indicates completely random prediction, while MCC=1 indicates perfect prediction. It is defined in equation (4):

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

Identification of novel putative anti-HIV AMPs from genome sequences

More than 1059 genome sequences were queried by the respective seven super-family profiles with the list of all proteome sequences (in the fasta format) searched retrieved, from the Ensembl database (http://www.ensembl.org/index.html) and the Uniprot database (http://www.uniprot.org/). A cut-off E-value was set to be 0.01 for the search of putative anti-HIV AMPs with the profiles built with HMMER. The "Query db" was used to identity peptides that may have the same signatures/motifs and properties as the profiles of the various super-families, thus the identified peptides were considered as putative anti-HIV AMPs and could have activity against HIV. This was accomplished using "hmmsearch" module of HMMER package and the command line employed is as follows:

hmmsearch–E 1e-2 family.hmm familyquery.txt>resultfile.txt … ……Where *family.hmm* in one of the seven super-families profiles, *familyquery.txt* represents the species scanned again the profile and *resultfile.txt* is the result file saved after querying that species against a particular super-family profile.

Physicochemical characterisation of the putative anti-HIV AMPs

The following physicochemical properties including: (*i*) the number of basic residues, (*ii*) acidic residues, (*iii*) net charge, (*iv*) the Isoelectric point, (*v*) the Boman Index (or protein binding potential), (*vi*) Hydrophobic residues, (*vii*) the instability index of the proteins [10,11], (*viii*) the number of Arginine (Arg) or Lysine (Lys) residues, (*ix*) the presence of Cysteine (Cys) residue [30] of the putative anti-HIV AMPs and HIV protein gp120 were calculated using the prediction interface of Bactibase [10,11] and APD [12,13]. This was accomplished with the amino acid sequences of the putative peptides and HIV protein gp120 as input in (http://bactibase.pfba-lab-tun.org/physicochem and http://aps.unmc.edu/AP/design/design_improve.php).

De novo structure predictions of the putative anti-HIV AMPs and HIV protein gp120

Prediction of the top 10 putative anti-HIV AMPs structures, based on their predictive E-values, as well as the structure of HIV gp120 protein were performed using I-TASSER (Iterative Threading ASSembly Refinement) server, which is an example of a *de novo* method of peptide or protein structure prediction [31]. In brief, the 3-D structures of the anti-HIV AMPs and gp120 protein were predicted by uploading each sequence onto the I-TASSER website. The user enters their email address to which the results link will be send. After, naming the uploaded sequence, the menu "Run I-TASSER" was selected [32]. The visualisations of the 3-D output structures were done using the PyMOL version 1.3.

Docking analysis of the putative anti-HIV AMPs and HIV proteins

The docking of the 10 putative anti-HIV AMPs to the HIV protein gp120 were accomplished using PatchDock Beta 1.3 version, a free online web-server that allows for protein-small ligand molecule docking, available at http://bioinfo3d.cs.tau.ac.il/PatchDock/ [33]. Briefly, the docking was done by uploading the respective PDB files from I-TASSER of gp120 protein and the putative anti-HIV AMPs onto the PatchDock server website, after which the user enters an email address. The cluster RMSD was set to 4.0 Å and the complex type was selected as "protein-small ligand". The task was submitted by selecting "Submit Form". The docking results were sent via an email notification, containing the web link to the docking results. Interaction analysis of the complex formation between the HIV protein and the putative anti-HIV AMP was done using PyMOL 1.3. Software.

Peptides synthesis

The selected putative anti-HIV AMPs were chemically synthesized by GL Biochem Ltd. (Shanghai 200241, China) using the solid-phase method and were purified to >98% by reverse-phase High-Pressure Liquid Chromatography and the AMPs were shipped in a lyopholilzed form.

Anti-HIV assays of the putative AMPs

The anti-HIV activity of the putative AMPs against HIV-1 pseudotyped virus-based assays was performed by the Biomed-Advanced Materials Division, Mintek (Pretoria, South Africa), as described in [34]. In brief, T cells were seeded the day before the antiviral testing at 3×10^5 cells/ml. The following day, the viability was checked and 2×10^5 cells/ml was placed into a 50 ml conical tube and HIV-1 NL4-3 stock added. The cells were incubated with the virus for 90 min. Cells were subsequently washed four times with 0.01 M PBS to remove any unbound virus. A control set of cells were incubated without the virus and washed four times with 0.01 M PBS to replicate the test cells. A total of 10 ml of 10% RPMI media was then added to

the cells and 100 µl of cells were added to each well of a Corning® Costar® 96-Well Cell Culture Plates (Sigma Aldrich, USA). The plate was placed into a 37°C, 5% CO_2 incubator to equilibrate for one hour. During the incubation, compounds were made up in 10% RPMI media containing 10% heat inactivated FCS. The compounds were made up to the desired concentration. A total of 100 µl of compound solution was added to the wells containing cells and mixed to ensure they were homogeneous. The plate was placed into a 37°C, 5% CO_2 incubator for five days. Following incubation, the microtiter plates were stained with XTT tetrazolium dye to evaluate the efficacy of the putative AMPs. The plates were then read on a multi-plate reader at 450 nm (xMARK™, Bio-Rad, USA) to determine the value of EC_{50} (50% inhibition of virus replication) of each AMP. A concentration of 50 µg/ml peptide was used in the screening process of the anti-HIV effect of the putative AMPs since the laboratory internal control achieves 50% HIV inhibition at this concentration, and subsequent anti-HIV activity of the peptides for the dose-dependent effect experiment was done with serial dilutions from 12.5 µg/ml to 150 µg/ml.

Cell viability assay

In vitro cytotoxicity test of putative anti-HIV peptides were performed by 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxy-phenyl)-2-(4-sulfophenyl)-2H-tetrazolium) (MTS) procedure [35]. Briefly, 100 µl of human T cell lines were seeded into 96-well plates at a density of 1×10^6 cells/ml the day of the experiment, and was incubated in 5% CO_2 atmosphere at 37°C, during which, test compounds were made up in 10% RPMI media to the desired concentrations. A total of 100 µl compound was added to the wells containing the cells and mixed to ensure the solution was homogeneous. The plate was placed into a humidified 37°C incubator with 5% CO_2 atmosphere for five days. Following 5 days incubation, 10 µl of MTS was added and mixed. The plate was further incubated for a further four hours at 37°C in a humidified, 5% CO_2 atmosphere, and the absorbance was recorded at 450 nm (xMARK™, Bio-Rad, USA).

Results

Curation of biologic datasets

Literature mining revealed that APD, CAMP, Cybase and UniProtKB had 79, 36, 23 and 27 experimentally validated anti-HIV peptides. Beside the experimentally validated anti-HIV AMPs, inferred anti-HIV and synthetic anti-HIV AMPs were also identified during this process (Supplementary Table 1). Further analysis was done to eliminate the duplicate anti-HIV AMPs found across the various databases, and a final list of 92 experimentally validated and seven predicted anti-HIV peptides were compiled. The list of anti-HIV AMPs was further classified into different families or super-families from which they originated, based on literature. This partitioning shows that 27, 4, 9, 5, 6, 8 and 33 anti-HIV AMPs were allocated to Amphibian, Microorganism, Human defensin, Fish and Crab, Insect, Vertebrate and Plant super-families respectively. However, only 4, 1 and 2 predicted anti-HIV peptides originated from Plants, Vertebrates and Microorganisms respectively (Supplementary Table 2).

Peptide data sets for model training and testing

The peptide sequences of the experimentally validated anti-HIV AMPs served as input data for the construction of super-family specific profiles. The selection of these sequences is justified since there are made up of amino acid residues which by the virtue of their presence within the sequence holds a direct implication for the functioning of the peptides, thus are considered motifs (identifiers) for model creation. Therefore, motifs generated by the profiles would help search for other peptide sequences with the same motifs and/or signatures. The sequences of these anti-HIV AMPs were randomly divided into the training set and the testing set for the purpose of model creation. The sequences of the anti-HIV AMPs used for both the training and the testing sets were partitioned at a 3: 1 ratio for each super-family (Supplementary Tables 3 and 4).

Independent testing of the seven profiles indicates their robustness

Following the design of the various profiles, each profile was tested against a blind data set, where each trained profile was tested against the testing data set of the same super-family from which the training data set was obtained for the building of its profile. Additionally, the trained profiles were scanned against a negative control data set, made up of 596 neuropeptides, having no recorded anti-HIV activity. The matches of the query profiles against training datasets are shown with scores (bits) and E-values. The E-value, which is calculated from the bits score, shows the number of true positives that are selected by the training dataset. Therefore, an E-value of 0.05 indicates that there is only 5% chance that the hit is false or has come up by chance. Hence, a low E-value is considered appropriate with the lowest E-value appearing at the top of the result list. A cut-off E-value was applied to the HMMER algorithm to strengthen the ability of the profile to discriminate between true positive anti-HIV AMP and false negative anti-HIV AMPs. In addition, it was possible to get the number of True positive (TP) AMPs from the total number of input sequences, thus the False positive (FN) could be extrapolated and the results are shown in Table 1, reflecting the capacity of each profile to distinguish a true anti-HIV AMP from a false anti-HIV AMP.

Performance measurement of the seven profiles indicates as highly specific, sensitive and accurate

After testing the ability of the trained profiles to recognise peptide sequences potentially exhibiting the same biological activity, and discriminate those with no anit-HIV activity, the performance was calculated with the aim to assess the robustness of each profile, by knowing their specificity, sensitivity, accuracy and MCC. Applying the formula stated in the method section, these parameters were determined as reported in Table 2.

Proteome sequence databases query and discovery of putative anti-HIV AMPs

The "Query db" stage to search for novel anti-HIV AMPs was to identity peptides that have the same signatures/motifs and properties as the created profiles of the various super-families. As elaborated at

Families or AMPs profile	True Positive (TP)	False Negative (FN)	True Negative (TN)	False Positive (FP)
Amphibians	3	2	577	19
Microorganisms	1	0	596	0
Human Defensins	2	0	587	9
Fish and Crabs	2	0	594	2
Insects	0	1	593	3
Vertebrates	1	1	590	6
Plants	6	5	565	31

Table 1: Results obtained by querying HMMER profiles against their respective testing dataset.

Super-families	Sensitivity (%)	Specificity (%)	Accuracy (%)	MCC
Amphibians	60	96.8	96.5	0.27
Microorganisms	100	100	100	1
Human Defensins	100	98.48	98.49	0.42
Fish and Crabs	100	99.66	99.66	0.71
Insects	0	99.5	99.33	-2.9e-3
Vertebrates	50	98.99	98.82	0.26
Plants	54.54	94.79	94.06	0.28

Table 2: Performance measurements generated for each super-family using the Model created by HMMER profile.

the "Profiles query" step with the testing datasets, the matches of the query profiles against the proteome sequences are also shown with scores (bits) and E-values. Contrary to the independent testing cut-off E-value of 5%, a cut-off E-value of 1% was applied at this step to search for putative AMPs.

After the scanning the proteome sequence databases to identify putative anti-HIV AMPs, observations were made that the peptides belonging to the Amphibian, Human Defensin, Insect and Plant super-families, were single domain peptides. However, the putative anti-HIV AMPs identified using the Vertebrate super-family profile, were multiple domain peptides, i.e., some parts (domains) of the proteins were predicted to have anti-HIV activity but not the entire protein sequence (Supplementary Figures 1-3). These peptides could not be considered as putative anti-HIV AMPs due to the fact that most active AMP sequences range from 10 to 100 amino acids in length (33). Additionally, the individual domains within the protein sequences of putative anti-HIV AMPs have value E-values higher than the cut-off set at 0.01. A final list of 30 AMPs was identified [Sequences not shown (PCT application Patent No. PCT/IB2015/058997) and (UK National application No. GB 1420695.7)]. The AMPs were ranked according to their *E*-values with those having the smallest *E*-values considered the most likely putative anti-HIV AMPs. There was a very low probability that these peptides were wrongly predicted to be anti-HIV AMPs. Thus only 10 AMPs having the lowest *E*-values were used for the *in-silico* interaction studies with HIV protein gp120 for the continuation of this particular study.

Physicochemical characterisation of putative anti-HIV AMPs and HIV protein gp120

The physicochemical properties of the 10 top-ranked putative anti-HIV AMPs were calculated as to ensure that these putative peptides shared similar features to all classes of AMPs, more specifically towards anti-HIV AMPs (Tables 3a and 3b). It is noted that all putative anti-HIV AMPs identified can be considered novel since none of the peptides matched any existing AMP in repository databases.

Prediction of the putative anti-HIV AMPs and HIV protein gp120 3-D structures

The output from I-TASSER server after predicting the AMPs and gp120 3-D structures included (Figure 1) an estimate of accuracy scoring of the predicted peptides and gp120 protein 3-D structures based on the C-score, TM-score and Root Means Square Deviation (RMSD). It was shown that the predicted 3-D structure of gp120 had a C-score of 2.00. Although all the putative anti-HIV AMPs gave C-score values higher than -1.5, molecule 1 however had a C-score of -1.83. It was reported that gp120 has a TM-score of 0.99; and whilst all putative AMPs had TM-scores above 0.5, Molecule 1 however had a TM-score of 0.49. On the other hand, the RMSD of all of the predicted structures had a value above 4 Å, except for molecule 1. Although the RMSD were not less that 1 Å, the topologies of the predicted 3-D structures are a consequence of it having a TM-score above 0.5, since there is a strong correlation between the RMSD and the TM-score of the predicted 3-D protein structure (Table 4) [32].

The C-score is a confidence score for estimating the quality of predicted models and ranges from -5 to 2. It is based on the relative clustering structural density and the consensus significance score of multiple threading templates used to estimate the accuracy of the I-TASSER predictions. A C-score cutoff > -1.5 indicates that the model has a correct fold. The TM-score is a scale for measuring the structural similarity between the predicted 3-D structure and the template structure. A TM-score >0.5 indicates a model of correct topology and a TM-score <0.17 means a random similarity. All HIV proteins have TM-scores greater than 0.5, meaning that the proteins have correct topology or structural shape [29]. The RMSD is the atomic deviation of the predicted molecules from the templates molecules used to predict their structures and a RMSD of 1Å is ideal for a good structure.

The images of the predicted AMPs are depicted as in Figure 1. Base on the C-score value of gp120, which achieved the maximal of 2.00 because the structure has been solved and deposited in the Protein Data Bank (PDB), we may therefore assume that the principle used for the structure prediction in this study was good. In addition, we might conclude that the statistical data from the putative AMP 3-D structures prediction were acceptable.

Protein-peptide interaction study of the putative anti-HIV AMPs bound to HIV protein gp120

The result output provides the highest geometric score of the complexes formed between HIV protein gp120 and putative anti-HIV AMP as a PDB file. Besides the geometric scoring system given in the result section of PatchDock, additional information includes the Atom Contact Energy (ACE), the area covered between the two molecules, the transformation coordinates during the molecular interaction and the PDB file of the complex formed as a ball and stick structure (Table 5) [33].

Even though all the putative anti-HIV AMPs showed a positive interaction with gp120 protein, the final list of AMPs that warranted further validation work were only selected if they bind gp120 protein, at the point of interaction with the CD4+ receptor. This approach is ideal since these peptides could form the backbone of lead compounds for entry inhibitor development for HIV therapeutics. Using this principle, only Molecule 1, 3, 8 and 10 were chosen as putative anti-HIV AMPs (Figure 2). However, Molecule 7 was also included to the list because it had a very high binding score to gp120.

The putative AMPs have anti-HIV activity

HIV type 1 based assays with MT-4 cells were implemented to confirm the anti-HIV effect of the molecules. The results revealed that Molecule 7, Molecule 8 and Molecule 10 could prevent the replication of NL4-3 virus in our preliminary experiments as compare to all other putative anti-HIV AMPs tested (Figure 3).

Subsequent dose-dependent experiments revealed that only Molecule 7 and Molecule 8 were able to significantly inhibit the HIV-1 NL4-3 replication as compared to Molecule 10, with Molecule 7 and Molecule 8 showing the highest anti-HIV percentage inhibition of HIV-1 NL4-3. It was possible to extrapolate from the dose-response curve that Molecule 7 and Molecule 8 have effective concentrations (EC) of 37.5 µg/ml and 93.75 µg/ml respectively. Hence, these molecules pave the way for the development of an entry inhibitory drug using the peptide as the lead compound. Consequently, Molecule 7 and Molecule 8 were selected to continue the peptides anti-HIV testing and to clarify their mechanisms and application. Conversely, the inhibitory ability of

	Mass	Most common amino acids and %	Lysine %	Arginine %	Cysteine %	Isoelectric point	Net charge	Total hydrophobic ratio	Protein-binding Potential (Boman Index)	Half Life in Mammals	Sequence similarity with other molecules and percentage
Molecule 1	8903.716 Da	Cys: 16	11.39	6.33	16	8.37	6	34 %	2.17 kcal/mol	1.2 h	SLPI: 68.22%
Molecule 2	4028.831 Da	Lys: 18.92	18.92	5.41	0.00	11.49	7	43 %	1.26 kcal/mol	1.3 h	Cecropin A: 86.48%
Molecule 3	4040.889 Da	Lys: 18.92	18.92	8.11	0.00	11.86	8	43 %	1.37 kcal/mol	1 h	Hyphancin IIIF: 81.08%
Molecule 4	4088.926 Da	Lys: 21.62	21.62	5.41	0.00	11.25	7	43 %	1.26 kcal/mol	1.3 h	Cecropin B: 83.78%
Molecule 5	4077.906 Da	Lys: 18.92	18.92	8.11	0.00	11.48	8	40 %	1.39 kcal/mol	1 h	Papiliocin: 76.92%
Molecule 6	4031.883 Da	Lys: 18.92	18.92	5.41	0.00	11.17	6	45 %	1.03 kcal/mol	1.3 h	Cecropin A: 72.97%
Molecule 7	4073.94 Da	Lys: 18.92	18.92	8.11	0.00	11.46	7	43 %	1.45 kcal/mol	1 h	Cecropin B: 94.59%
Molecule 8	3670.552 Da	Cys: 17.65	14.71	5.88	17.65	9.60	8	38 %	1.07 kcal/mol	1.2 h	hBD2: 82.92%
Molecule 9	2780.401 Da	Cys: 22.22	7.41	0.00	22.22	6.03	0	51 %	-0.11 kcal/mol	7.2 h	Cliotide T1: 76.66%
Molecule 10	3908.564 Da	Ala: 16.67	11.11	5.56	0.00	10.33	2	47 %	1.33 kcal/mol	2.8 h	Cecropin D: 80.55%
Kn2-7	1674.152 Da	Arg & Ile: 23	15	23	0.00	12.81	5	61 %	1.8 kcal/mol	1.1 h	Bmkn2: 62.5%
MucroporinS1	1091.39 Da	Ser & Leu: 18	9	0	0.00	9.70	1	54 %	-1.19 kcal/mol	1.9 h	Mucroporin: 64.7%

Table 3a: Physicochemical properties and parameters for the 10 putative anti-HIV AMPs, the positive control (Kn2-7) and the negative control (Mucroporin-S1).

HIV protein	Mass	Isoelectric point	Net charge	Total hydrophobic ratio	Instability Index	Half Life in Mammals
gp120	35098.29 Da	7.52	+6	37%	41.74	1 h

Table 3b: Characterisation of the different physicochemical properties and parameters for HIV protein gp120.

Putative anti-HIV AMPs	C-score	Exp. TM score	Exp. RMSD (Å)
Molecule 1	-1.83	0.49 ± 0.15	7.3 ± 4.2
Molecule 2	-0.06	0.71 ± 0.12	2.2 ± 1.7
Molecule 3	0.03	0.72 ± 0.11	2.0 ± 1.6
Molecule 4	-0.05	0.71 ± 0.12	2.2 ± 1.7
Molecule 5	-0.01	0.71 ± 0.11	2.1 ± 1.7
Molecule 6	0.04	0.72 ± 0.11	2.0 ± 1.6
Molecule 7	-0.00	0.72 ± 0.11	2.1 ± 1.7
Molecule 8	0.95	0.84 ± 0.08	0.5 ± 0.5
Molecule 9	0.73	0.81 ± 0.09	0.5 ± 0.5
Molecule 10	0.06	0.72 ± 0.11	2.2 ± 1.7
Positive control (Kn2-7)	0.14	0.73 ± 0.11	0.5 ± 0.5
Negative control (Mucroporin-S1)	0.28	0.75 ± 0.10	0.5 ± 0.5
gp120	2.00	0.99 ± 0.03	1.7 ± 1.5

Table 4: Quality assessment scores of the predicted 3-D structures of the putative anti-HIV AMPs, the positive and negative controls.

Molecule 10 was not convincing since concentration up to 150 µg/ml was unable to inhibit more than 30% of the virus (Figure 4).

The putative anti-HIV AMPs has a selective effect to HIV-1 NL4-3

To confirm the ability of the putative AMPs to exhibit anti-HIV activity, we carried out a non-selective cytotoxicity assay of Molecule 7, Molecule 8 and Molecule 10 on MT-4 cell lines to establish the selective potential of these compounds. The assay was performed by treating the T cell lines with different concentrations of the compounds for 5 days and measuring the viability of the cells by taking the absorbance at 450 nm, on an ELISA plate reader (xMARK™, Bio-Rad, USA). The "Statistical Package for the Social Sciences" (SPSS) was used to calculate the Cytotoxic Concentration (CC) values and Figure 5 showed the CC of the three molecules achieved with various doses of the peptides. Even thought the CC_{50} was not established, it should be noted that at 150 µg/ml 80% of T cells was still not inhibited by any of the peptides. Therefore, increasing the peptides concentration would help determine the CC_{50} of each compound.

Discussion

Mining of data sets

The use of AMPs as alternative sources for drug design has encouraged a massive explosion in the research area of these biomolecules. AMPs exhibit certain characteristics and properties namely net positive charges, hydrophobicity, high specificity towards microorganisms and low microbial resistance that is exploited in research and their use as novel drug compounds. Many AMPs have been proven to have activity towards various gram-positive and gram-negative bacteria, protozoa, cancer, fungi as well as viruses [14,15,36-38]. Some AMPs have even shown potential anti-HIV activity [9].

Different molecular approaches have been used to identify, comprehend their mechanism of action and validate the activity of these peptides. The race and progression of scientific technologies have brought about alternative methods to elucidate AMP functions in addition to molecular techniques, hence the implementation of computational biology [39,40]. These computational approaches have facilitated easy identification of AMPs because they are less time consuming, less laborious and affordable. Many experimentally validated anti-HIV AMPs have been deposited into curated repository databases [10-16]. Due to the fact that the purpose of data mining was to retrieve experimentally validated anti-HIV AMP sequences, which aided in the building of the HMMER profiles/models, these repository databases were the suitable choice since they harbour curated AMPs, whose anti-HIV activity have been proven. The AMPs specific repository databases reduced the time to search for anti-HIV AMPs since it have been arranged in a user friendly manner and are well curated to contain AMPs with various activities: anti-cancer, anti-bacterial, anti-parasite as well as anti-HIV.

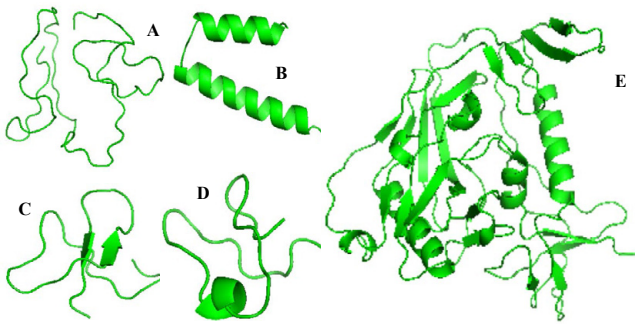

Figure 1: Three-Dimensional structure of some putative AMPs and HIV protein gp120 predicted using I-TASSER. A: Molecule 1, B: Molecule 2, C: Molecule 8, D: Molecule 9, E: gp120 protein. The AMPs showed different secondary structures including α-helices, β-sheets and extended shapes. Molecule 3, Molecule 4, Molecule 5, Molecule 6, Molecule 7 and Molecule 10 all have the similar structre as Molecule 2.

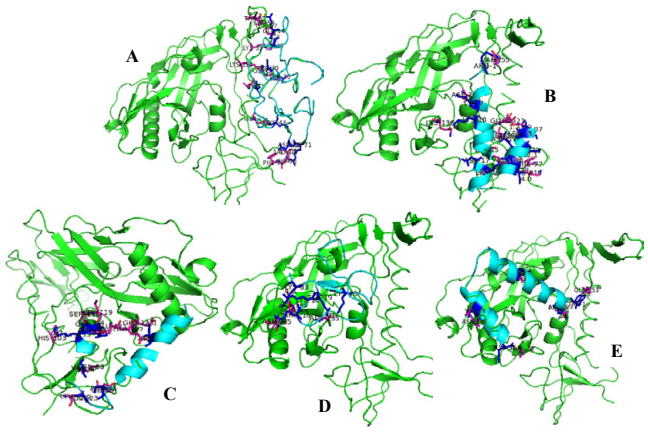

Figure 2: Interaction of HIV-1 protein gp120 and the putative anti-HIV AMPs. A: Molecule 1-gp120; Molecule 3-gp120; C: Molecule 7-gp120; D: Molecule 8-gp120; E: Molecule 10-gp120. The putative AMPs bind to gp120, at the area where CD4+ molecules interact with gp120. The cartoon representation in green colour is the HIV protein gp120 and the putative anti-HIV AMP is represented in light blue colour. The purple colour represents the stick representation of gp120 amino acids interacting with the amino acid of the putative AMPS, represented with a dark blue stick.

Literature mining of datasets

To create highly discriminatory profiles/models, it had to be ensured that the AMPs retrieved from the repository databases were thoroughly curated and had the intended anti-HIV activity to discover novel members of this peptide class. The activity was verified through literature mining after retrieving the relevant publication of each AMP from the curated databases. This is an essential step as it aids in the construction of a profile of a particular family or super-family with specific activity.

Models testing

After building the profiles, the aim was to ensure that the models created were indeed robust and would identify peptides sequences with anti-HIV activity; hence the partitioning of 25% of the experimentally validated AMPs, from the initial list of retrieved anti-HIV AMPs to scan the built profiles. Although the total number of true positives could not be obtained for all the built profiles, the number of false negatives for each profile was still smaller when compared to that of true positives except for the "insects" profile which testing set was one. This could be due to the small size of the training sets used in model creation. However, the scanning of the profiles against non-anti-HIV AMPs showed that the number of true negatives for each profiles was far higher than the false positives (Table 1). The results obtained from scanning the testing sets against the built profiles confirmed the robustness of the profiles.

Models evaluation

The performance evaluation of each model was necessary to ensure correctness of each created profiles. The performance indications of the respective models achieved by calculating the sensitivity, accuracy, specificity and the MCC were observed for the seven super-families. The sensitivity score of 60%, 100%, 100%, 100%, 0%, 50% and 54.54% were obtained for the Amphibian, Microorganism, Human defensin, Fish and

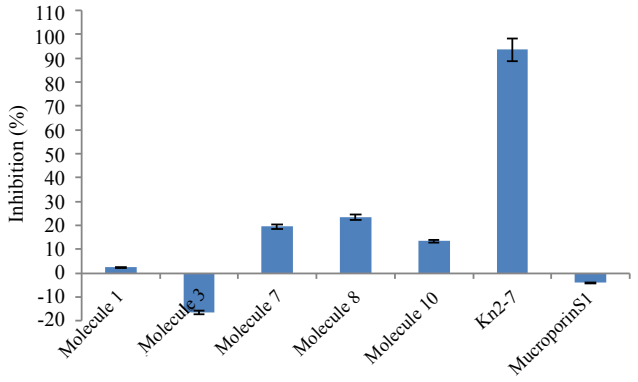

Figure 3: Screening of the putative AMPs against HIV-1 NL4-3 using a single dose. The inhibition of HIV-1 NL4-3 by the putative AMPs was measured by using 50 µg/ml. Only Molecule 7, Molecule 8 and Molecule 10 showed some inhibition of the virus replication, by preventing their entry into the host cell to multiply. Thus molecules 7 and 8 block the binding of HIV gp120 to CD4+ of T cells to allow the virus replication. At this concentration, the Kn2-7 (positive control) showed approximately 93.63% inhibition whilst Mucroporin-S1 (negative control) showed no inhibition of the virus. All the data represent the mean values for three independent experiments and are reported as mean ± SD of the three replicated samples of each Molecule.

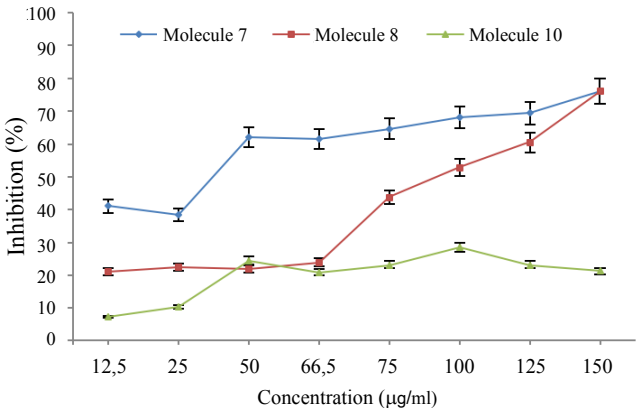

Figure 4: Dose-dependent effects of Molecule 7, 8 and 10 against HIV-1 NL4-3. The EC_{50} of Molecule 7 was 37.5 µg/ml and that of Molecule 8 was 93.75 µg/ml. We could not calculate the EC_{50} of Molecule 10 due to the fact that even at a concentration of 150 µg/ml it did not inhibit 30% of the virus. All the data represent the mean values for six independent experiments and are reported as mean ± SD of the six replicated samples of each molecule.

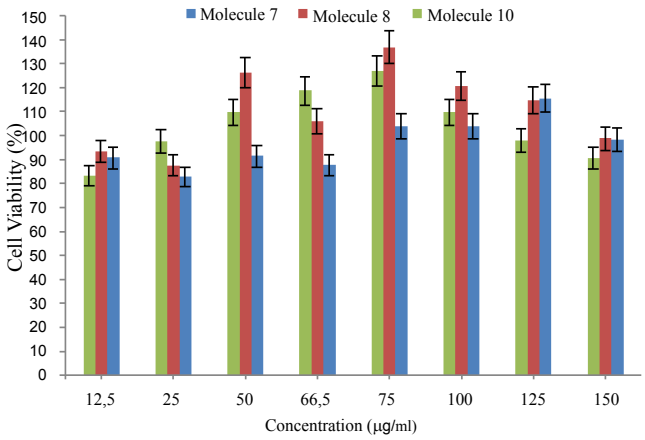

Figure 5: Cytotoxicity of Molecule 7, 8 and 10 on T cell lines. All the data represent the mean values for six independent experiments and are reported as mean ± SD of the six replicated samples of each molecule.

	Binding affinity geometric scores
Putative AMPs	**gp120**
Molecule1	14926
Molecule2	12650
Molecule3	13686
Molecule4	11968
Molecule5	12708
Molecule6	13104
Molecule7	13648
Molecule8	11086
Molecule9	9186
Molecule10	12208
Positive control (Kn2-7)	8158
Negative control (Mucroporin-S1)	6912

Table 5: Geometric scores of the binding affinity obtained from the docking of the putative anti-HIV AMPs and HIV-1 protein gp120.

Crab, Insect, Vertebrate and Plant super-families respectively models. Although it is said that the calibration step in HMMER algorithm increases the sensitivity of the models during the testing of the models and the genome/proteome sequence scanning of databases [40,41], the performance of the calibrated profile is only rated on its accuracy and specificity performances. All the models had accuracies and specificities above 95% except for the Plants super-family model, of which the accuracy and the specificity were 94.79% and 94.06% respectively. This proves that the created profiles had more than 95% confidence to predict a peptide as a putative anti-HIV AMP. Whilst the MCC ranges from -1 to +1, MCC equal to zero indicates a completely random prediction, MCC equal to 1 indicates perfect prediction and MCC equal to -1 indicated a total disagreement between the prediction and the observation [42]. It was also observed that the MCC of the individual super-family models constructed varied between the range of -1 and 1, with the Microorganisms model having a MCC of 1 since the profile scored 100% for the other parameters evaluated. However, the MCC of the Insects model was negative due of the fact the model could not identify its testing set and had a sensitivity of 0%. Nonetheless, the specificity and the accuracy proved that the model was of good quality as both parameters had 99.5% and 99.33% validation respectively (Table 2).

Discovery of novel putative anti-HIV AMPs

The models were scanned against various proteome sequence databases to identify putative AMPs, which may have HIV activity. The results of the database query steps also called "Query db", resulted in a ranked list of the best scoring domains in the order of occurrence of the sequence and alignments for the highest scoring domains. The high conserved residues for both the query sequence and the consensus pattern of the profile of a super-family are shown in capital letters for the results. It was observed that the numbers of putative AMPs predicted to have anti-HIV activity by the algorithm were not randomly selected as the high sensitivity of the HMMER profiles was due to the combination of the scoring system and the low cut-off *E*-value calculated during the proteome sequences scanning steps. The *E*-value gives more information about the probability of that predicted AMPs to be true positive or false negative anti-HIV AMPs [17,18,22]. The list of the putative anti-HIV AMPs generated by HMMER super-families specific profiles truly indicated the diversity of the Antimicrobial Peptides as the predicted AMPs were identified from different species. Whilst the best anti-HIV AMPs had an *E*-value of 1.4e-54, meaning that there is only a 1.4e-54% chance for the peptide to be a false predicted anti-HIV AMP, the lowest score observed has an E-value of 9.4e-3, meaning that there is only a 9.4e-3% chance for the peptide to be predicted as a false anti-HIV AMP. Hence, all the putative anti-HIV AMPs predicted by HMMER had higher probability scores to be considered true anti-HIV AMPs.

Putative anti-HIV AMPs physicochemical properties and structures prediction

Although the study was able to identify 30 AMPs, only the 10 best putative anti-HIV AMPs with the lowest E-values were used for the continuation of the study based on their ranking. The 10 best putative anti-HIV AMPs were firstly characterized to ascertain their physicochemical properties before subsequent usage in the docking studies. The observation was made that none of the AMPs had 100% similarity to the existing experimentally validated AMPs or anti-HIV AMPs, thus the novelty of the sequences. Nine out of the 10 putative anti-HIV AMPs had a net positive charge. However, only Molecule 9 had a zero net charge and this was contributed to by the low percentage of the presence of the positively charged amino acids Lysine and Arginine within the AMP sequence. These amino acids in addition to Histidine are the primary biomolecule residues, which trigger the electrostatic interaction of the AMP to the pathogen receptor [43]. The zero net charge of Molecule 9 was shown to have a direct affect on its protein binding potential (Boman index) and its binding capacity to gp120 protein since it was lower than zero. An AMP binding potential value ≤ 1 indicates that the peptide will likely have increased antimicrobial activity without many side effects. AMPs with Boman indices less than zero have been shown to only have antibacterial activity, and AMPs with a higher index value (2.50-3.00) indicates that the peptide is multifunctional with hormone-like activities [44]. It is a well-known fact that the most potent anti-HIV AMPs have a high presence of Cysteine residues [25]; only a few AMPs (Molecule 1, 8 and 9) were shown to have a percentage of Cysteine higher than 16% (Table 3a). This amino acid residue may be a contributor for a proper folding of the peptide, in order for its to fit at the binding point where gp120 interacts with CD4+ T cells. The hydrophobic content of the individual putative anti-HIV AMP was all above 30%, which is the expected value of Antimicrobial Peptide hydrophobicity. This will contribute to the affinity binding of the putative anti-HIV AMPs to the virus and destruction of the viral membrane through the many mechanisms employed by AMPs to interact and subsequently destroy their targets [42,45].

Modeling of putative AMPs and gp120 protein

The different parameters (C-score, TM-score, RMSD) used to evaluate the prediction of the putative AMPs and gp120 protein 3-D structures enable the validation of the structure modeling. The C-score of all the predicted 3-D structures of Molecules 2-10 and gp120 protein were above the limiting value of -1.5 especially the C-score of gp120, for which the 3-D structure has already been solved. This lends credibility to the use of this *In Silico* tool for the prediction of 3-D structures. The C-scores observed for the molecules could be as a result of the fact that there are available or correct templates to base their modeling and prediction of the structures on. Conversely, C-score of Molecule 1 was below the limit of -1.5 and was reported to be -1.83; and could indicate that the molecule did not have an available template for its modeling [32]. Similar to the C-score, the TM-score of the predicted molecules were also higher than the cut-off value of 0.5, except for Molecule 1, which TM-score was 0.49. Their TM-score being higher than 0.5 signified that Molecules 2-10, HIV protein gp120, the positive and the negative controls have structural similarity with the templates that were used to predict their structures [32,46]. The result of Molecule 1 does not always achieve the threshold value imposed by each parameter of I-TASSER, thus it can be concluded that there is a strong correlation between the Root Mean Square Deviation (RMSD) and the TM-score of a predicted protein or peptide [46]. Although there is not a defined RMSD value for 3-D structure prediction, a RMSD value of 2-4 Å is considered good and a RMSD ≤ 1 Å is considered ideal. Thus, Molecules 2-10 having a RMSD within the accepted range had less distance and atomic deviation between the superimposed peptides and the templates, which were used for their 3-D structure prediction [47-49]. Furthermore, the 10 putative anti-HIV AMPs exhibit various secondary structures including α-helical, parallel β-sheet, anti-parallel β-sheet, the extended and the loop conformational structures. This is in line with the various structural conformations exhibited by AMPs in as well as AMPs considered to be anti-HIV AMPs specifically.

Putative AMPs interaction with gp120

The interaction of the putative AMPs to gp120 protein was to confirm the ability of these peptides to bind the protein, at the right conformation, so as to prevent the interaction of gp120 with CD4+ T cells in an infected individual. The binding affinity of the putative anti-HIV AMPs corroborates the hypothesis that AMPs with a net positive charge less than +2 would have less electrostatic attraction towards the pathogenic organisms, thus less binding affinity to these organisms. Molecule 9 which had a zero net charge, presented the lowest binding affinity interaction with the HIV protein gp120. The zero net charge of this molecule was shown to have the lowest Boman index as well, which was confirmed with the docking study. Hence, the deduction could be made that the positive charges of an AMP influences the binding of that peptide to its target [42,45]. However, Molecule 1 which showed the highest binding affinity to gp120 can be contributed to by its positive net charge of + 6, a good Boman index, the presence of Cysteine amino acids and the fact that it had the lowest E-value. Conversely, the binding affinities between the 10 putative AMPs and gp120 did not show a parallel decrease from the AMPs with the highest E-value (most probable to be an anti-HIV AMP) to the putative anti-HIV AMP with the lowest E-value (less probable to be an anti-HIV AMP). The observation might be contributed by the various parameters that contribute to a good anti-HIV peptide all of which were not taken into consideration by this study during model prediction.

The invasion of macrophages/monocytes, T lymphocytes, by HIV is through the contact of CD4+ surface molecules of these cells with HIV surface protein gp120 [50-52]. From this perspective, it would justify the use of Molecules 1, 3, 8 and 10, as putative anti-HIV AMPs to test for anti-HIV activity against HIV since these AMPs bind specifically to gp120, at the point of interaction where gp120 protein interacts with CD4+ T cells or macrophages/monocytes. Therefore, these peptides are of great interest and their binding to gp120 could be a new strategy to target the HIV-1 and counteract its entry into T cells and thus could prevent the virus propagation through the contamination of the T cells population by competing for this target with HIV [51]. The addition of Molecule 7 to the list of potential anti-HIV AMPs was due to the fact that it had a high binding score to gp120, and binds at the areas where DC-SIGN molecules of dendritic cells (DCs) interact with gp120 [53].

Anti-HIV activity and cytotoxicity

The validation of the five putative anti-HIV AMPs reveals that after the screening process of the compounds against HIV-1 NL4-3 virus, only Molecule 7, 8 and 10 could inhibit HIV with 50 µg/ml of each peptide (Figure 3). The dose-dependent response assays aided in establishing the EC_{50} of Molecule 7 and Molecule 8 which showed that these peptides inhibited HIV-1 NL4-3, with EC_{50} values of 37.5 µg/ml and 93.75 µg/ml respectively (Figure 4). The assays further prove that the HMMER algorithm could be used for the design and discovery of putative AMPs, to target a particular disease, taking into account the previous experimentally validated active compounds that would enable the construction profiles to search for novel AMPs. Even thought the EC_{50} seem to be of high dosage to inhibit the viral growth, it has to be taken into account that various cytotoxicity concentrations of the Molecule 7 and Molecule 8 could still not inhibit 80% of T cells, for up to a concentration of 150 µg/ml (Figure 5), thus these peptides did not affect the T cells immunological function in the infected sample. These AMPs rather prevent the attachment of HIV gp120 to CD4+ of T cells, without affecting the T cells natural role but rather have a selective toxicity toward HIV. However, establishing the CC_{50}, the therapeutic index (TI) or selective index ($SI=CC_{50}/EC_{50}$) of each AMP could be within an acceptable range [54]. The therapeutic ability of Molecule 7 and Molecule 8 would inhibit the viral replication and could prevent the invasion of new T cells by directly blocking gp120 contact with the CD4+ of T cells, macrophages/monocytes [50,51]; DC-SIGN surface protein of dendritic cells [53].

Conclusion

The fight against the HIV pandemic to reduce the progress of the disease and to search for adequate therapeutic compounds is still an active area of research. Increased scientific research has enabled the design of many drugs that attack the virus at various steps of its entrance into human cells and/or during the virus life cycle. However, combined methods and therapies have helped reduce HIV replication and thus the mortality rate. One example of an anti-HIV drug, which prohibits the virus entrance into the human cells, is the new class of HAART, which act as entry inhibitors or fusion inhibitors. Enfuvirtide is the only anti-HIV peptide-based drug of this class of HAART, which has received approval from the FDA [55], and although other peptide-based drugs are either FDA approved [56] or are under clinical trials [55], more AMPs ought to be screened to develop potent anti-HIV entry inhibitors treatment regimens. The use of the HMMER algorithm is deemed an appropriate tool, which enables a more sophisticated search for novel peptides through the proteome sequence scanning. Several of the AMPs discovered through this study, have been proven to detect the HIV capsid protein p24 and a diagnostic kit has been developed employing these AMPs instead of the classically used antibodies [57].

Additionally, the ability of these putative AMPs to inhibit HIV-1 NL4-3 has been preliminary proven in this study, and it can be speculated that this activity is due to the fact these putative AMPs bind gp120 at the area where gp120 protein binds to CD4+ T cells, macrophages/monocytes, as shown in the *In Silico* docking study of gp120 and the identified AMPs. Taking this in consideration, the interaction of HIV to CD4+ molecules can potentially be prevented, and the result observed that a few of these peptides could inhibit HIV-1 NL4-3, reinforce their application for the development of peptide-based drugs with these molecules as lead compound.

Future Works

Future work will include the *In Silico* site directed mutagenesis of amino acids to optimize these peptides as to have a more potent effect to prevent the binding of HIV-1 gp120 protein to CD4+ molecules. This would be followed by an *in vitro* study of the anti-HIV activity of the mutated AMPs. In additional, the CC_{50} of all AMPs will be determined as well as specify the therapeutic index or selective index of the optimized AMPs. The anti-HIV activity of these AMPs will also be carried out on different HIV-1 pseudotype viruses to determine their broad activity. Furthermore, various assays will help to determine the application of the molecules as preventive/prophylactic drugs or therapeutic drugs. Finally, the complex formed between gp120 and putative AMPs will be solved using structural biology to validate the observations made by the *In Silico* binding study.

Acknowledgement

The authors of this study would like to acknowledge all funding contributors, the National Research Foundation, Medical Research Council, Department of Science and technology and the MINTEK/Nanotechnology Innovation Centre. Also special thanks to the Biomed-Advanced Materials Division, Mintek (Pretoria, South Africa), for their facilitation of part of this project.

Author's Contribution

Conceived and designed the experiments: AP MBT MNG. Performed the experiments: MBT MNG. Analyzed the data: AP MBT MNG. Contributed reagents/materials/analysis tools: AP MBT MNG. Wrote the paper: AP MBT MNG.

Funding Contributors

National Research Foundation, Department of Science and Technology/Mintek, Nanotechnology Innovation Centre and the Medical Research Council.

References

1. Gallo RC (2006) A reflection on HIV/AIDS research after 25 years. Retrovirology 3: 72.
2. Jenssen H, Hamill P, Hancock RE (2006) Peptide antimicrobial agents. Clin Microbiol Rev 19: 491-511.
3. Hancock REW, Lehrer R (1998) Cationic peptides: A new source of antibiotics. Trends Biotechnol 16: 82-88.
4. Chang TL, François F, Mosoian A, Klotman ME (2003) CAF-mediated human immunodeficiency virus (HIV) type 1 transcriptional inhibition is distinct from alpha-defensin-1 HIV inhibition. J Virol 77 : 6777-6784.
5. Wang W, Owen SM, Rudolph DL, Cole AM, Hong T, et al. (2004) Activity of α- and θ-defensins against primary isolates of HIV-1. J Immunol 173: 515-520.
6. Wu Z, Cocchi F, Gentles D, Ericksen B, Lubkowski J, et al. (2005) Human neutrophil alpha-defensin 4 inhibits HIV-1 infection *in vitro*. FEBS Lett 579: 162-166.
7. Chang TL, Klotman ME (2004) Defensins: Natural Anti-HIV peptides. AIDS Reviews 6: 161-168.
8. Münk C, Wei G, Yang OO, Waring AJ, Wang W, et al. (2003) The theta-defensin, retrocyclin, inhibits HIV-1 entry. AIDS Res Hum Retrov 19: 875-881.
9. Chen Y, Cao L, Zhong M, Zhang Y, Han C, et al. (2012) Anti-HIV-1 Activity of a new scorpion venom peptide derivative Kn2-7. PLoS ONE 7: e34947.
10. Hammami R, Zouhir A, Hamida BJ, Fliss I (2007) BACTIBASE: A new web-accessible database for bacteriocin characterization. BMC Microbiol 7: 89.
11. Hammami R, Zouhir A, Le Lay C, Hamida BJ, Fliss I (2010) BACTIBASE second release: A database and tool platform for bacteriocin characterization. BMC Microbiol 10: 22.
12. Wang Z, Wang G (2004) APD: The antimicrobial peptide database. Nucleic Acids Res 32: D590-D592.
13. Wang G, Li X, Wang Z (2009) APD2: The updated antimicrobial peptide database and its application in peptide design. Nucleic Acids Res 37: D933-937.
14. Fjell CD, Hancock REW, Cherkasov A (2007) AMPer: A database and an automated discovery tool for antimicrobial peptides. Bioinformatics 23: 1148-1155.
15. Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S (2010) CAMP: A useful resource for research on antimicrobial peptides. Nucleic Acids Res 38: D774-D780.
16. Sundararajan VS, Gabere MN, Pretorius A, Adam S, Christoffels A, et al. (2011) DAMPD: A manually curated antimicrobial peptide database. Nucleic Acid Res.
17. Brahmachary M, Krishnan SPT, Koh JLY, Khan AM, Seah SH, et al. (2004) ANTIMIC: A database of antimicrobial sequences. Nucleic Acids Res 32: 586-589.
18. Fjell CD, Jenssen H, Hilpert K, Cheung WA, Panté N, et al. (2009) Identification of novel antibacterial peptides by chemoinformatics and machine learning. J Med Chemother 52: 2006-2015.
19. Torrent M, Di Tommaso P, Pulido D, Nogués MV, Notredame C, et al. (2012) AMPA: An automated web server for prediction of protein antimicrobial regions. Bioinformatics 28: 130-131.
20. Juretić D, Vukicević D, Petrov D, Novković M, Bojović V, et al. (2011) Knowledge-based computational methods for identifying or designing novel, non-homologous antimicrobial peptides. Eur Biophys J 40: 371-385.
21. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14: 755-763.
22. Boman HG (1995) Peptide antibiotics and their role in innate immunity. Annu Rev Immunol 13: 61-92.
23. Mulvenna JP, Wang C, Craik DJ (2006) CyBase: A database of cyclic protein sequence and structure. Nucleic Acids Res 34: 192-194.
24. Wang CK, Kaas Q, Chiche L, Craik DJ (2008) CyBase: A database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Res 36: C206-C210.
25. Uniprot C (2009) The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 38: D142-D148.
26. Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659.
27. Gennaro R, Zanetti M (2000) Structural features and biological activities of the cathelicidin-derived antimicrobial peptides. Biopolymers 55: 31-49.
28. Vizioli J, Salzet M (2002) Antimicrobial peptides from animals: Focus on invertebrates. Trends Pharmacol Sci 23: 494-496.
29. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22 : 4673-4680.
30. Wang G, Buckheit KW, Mishra B, Lushnikova T, Buckheit RW (2011) *De novo* design of antiviral and antibacterial peptides with varying loop structures. J AIDS Clinic Res S2: 003.
31. Eswar N, Sali A (2008) Protein structure modelling. from molecules to medicines 139-151.
32. Roy A, Kucukural A, Zhang Y (2010) I-TASSER a unified platform for automated protein structure and function prediction. Nature Protocols 5: 725-738.
33. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. Nucleic Acids Res 33: W363-W367.

34. Montefiori DC (2005) Evaluating neutralizing antibodies against HIV, SIV and SHIV in luciferase reporter gene assays. Curr Protoc Immunol 12: 12.11.

35. Cory A, Owen T, Barltrop J, Cory J (1991) Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. Cancer communications 3: 207.

36. Wang CK, Colgrave ML, Gustafson KR, Ireland DC, Goransson U, et al. (2008) Anti-HIV cyclotides from the Chinese medicinal herb *Viola yedoensis*. J Nat Prod 71: 47-52.

37. Wang G, Watson KM, Buckheit RW (2008) Anti-human immunodeficiency virus type 1 activities of antimicrobial peptides derived from human and bovine cathelicidins. Antimicrob Agents Chemother 52: 3438-3440.

38. Hogeweg P (2011) The roots of bioinformatics in theoretical biology. PLoS Comput Biol 7.

39. Ouzounis CA (2012) Rise and demise of bioinformatics? Promise and progress. PLoS Comput Biol 8: e1002487.

40. Eddy S (2003) HMMER User's guide: Biological sequence analysis using profile hidden Markov models. Washington University School of Medicine 19: 65-85.

41. Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16: 412-424.

42. Hancock RE, Sahl HG (2006) Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat Biotechnol 24: 1551-1557.

43. Hancock REW, Chapple DS (1999) Peptide antibiotics. Antimicrob Agents Chemother 43: 1317-1323.

44. Boman HG (2003) Antibacterial peptides: Basic facts and emerging concepts. J Intern Med 254: 197-215.

45. Giuliani A, Pirri G, Nicoletto SF (2007) Antimicrobial peptides: An overview of a promising class of therapeutics. CEJB 2: 1-33.

46. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.

47. Park BH, Huang ES, Levitt M (1997) Factors affecting the ability of energy functions to discriminate correct from incorrect folds. J Mol Biol 266: 831-846.

48. Wei L, Huang ES, Altman RB (1999) Are predicted structures good enough to preserve functional sites? Structure 7: 643-650.

49. Carugo O, Pongor S (2001) A normalized root-mean-square distance for comparing protein three-dimensional structures. Protein Science 10: 1470-1473.

50. Chan DC, Fass D, Berger JM, Kim PS (1997) Core structure of gp41 from the HIV envelope glycoprotein. Cell 89: 263-273.

51. Kowalski ML, Potz J, Basiripour L, Dorfman T, Goh WC, et al (1987) Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. Science 237: 1351-1355.

52. Lu M, Blackow S, Kim P (1995) A trimeric structural domain of the HIV-1 transmembrane glycoprotein. Nature Struct Biol 2: 1075-1082.

53. Gringhuis SI, van der Vlist M, van den Berg LM, den Dunnen J, Litjens M, et al. (2010) HIV-1 exploits innate signaling by TLR8 and DC-SIGN for productive infection of dendritic cells. Nature Immunol 11: 419-426.

54. Becker DE (2007) Drug therapy in dental practice: General principles Part-2 pharmacodynamic considerations. Anesth Prog 54: 19-24.

55. Dwyer JJ, Wilson KL, Davison DK, Freel SA, Seedorff JE, et al. (2007) Design of helical, oligomeric HIV-1 fusion inhibitor peptides with potent activity against enfuvirtide-resistant virus. Proc Natl Acad Sci USA 104: 12772-12777.

56. Kilby JM, Lalezari JP, Eron JJ, Carlson M, Cohen C, et al. (2002) The safety, plasma pharmacokinetics and antiviral activity of subcutaneous enfuvirtide (T-20), a peptide inhibitor of gp41-mediated virus fusion, in HIV-infected adults. AIDS Res Hum Retroviruses 18: 685-693.

57. Williams ME, Tincho MB, Gabere NM, Uys A, Meyer M, et al. (2016) Molecular validation of putative antimicrobial peptides for improved Human Immunodeficiency Virus diagnostics via HIV protein p24. J AIDS Clinic Res 7: 571.