2-1-2016

In silico screening, analysis, and modelling for a novel anticancer peptide

Youssef Abdou

THE AMERICAN UNIVERSITY IN CAIRO
الجامعة الأمريكية بالقاهرة

School of Science and Engineering

In silico screening, analysis, and modelling for a Novel anticancer peptide

A Thesis Submitted to

Biotechnology

in partial fulfillment of the requirements for
the degree of Master of Science

by Youssef Abdou

(under the supervision of Dr. Asma Amleh)
December 2016

THESIS PROPOSAL APPROVAL

Student Full Name: _____

Student ID: _____

Thesis Title: _____

Please attach the thesis proposal to this form.

Proposed Graduation Date: _____

Thesis title and proposal have been approved:

Thesis Supervisor:
Name: _____
Title: _____
Department: _____
Institution: _____
Signature: _____Date: _____

First reader:
Name: _____
Title: _____
Department: _____
Institution: _____
Signature: _____Date: _____

Second reader:
Name: _____
Title: _____
Department: _____
Institution: _____
Signature: _____Date: _____

Department Chair:
Name: _____
Title: _____
Signature: _____Date: _____

Student:
I, (Student Name) _____
have read and understood the thesis guidelines.
Signature: _____Date: _____

DEDICATION

I dedicate this work to my parents, Dalal Kamal Aly Sabet and Tarek Said Abdou, and my brother, Sherif Tarek Said Abdou; thank you for all your love, support, and motivation.

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor Dr. Asma Amleh for her endless support, time, effort, and guidance.

I would also like to Acknowledge Mr. Mustafa Adel for his contributions to our project. Finally, I would like to Acknowledge Mr. Amged Ouf for his valuable advice and support.

The American University in Cairo

In silico screening, analysis, and modelling for a Novel anticancer peptide

Youssef Tarek Abdou

Dr. Asma Amleh

ABSTRACT

Cancer is currently one of the leading causes of mortality and morbidity worldwide. Most anticancer therapies rely on small molecule drugs (<0.5 kDa). As with all small molecule drugs, chemotherapy is highly toxic and presents many off-target side effects. Peptide drugs offer improved specificity and are cheaper and more accessible to manufacture. In this study, we have developed a support vector machine (SVM) model in order to detect peptide sequences with potential anticancer activity through scanning the Red Sea Metagenomic library. Furthermore, we conducted an *in silico* study in order to analyze one of the peptides returned by the SVM pipeline and assessed its cytotoxicity and the mode of cell death by conducting MTT and Annexin V staining assays, respectively. We observed that the selected anticancer peptide contains the C-terminal portion of the homeodomain structure, of human Pax6, an antennapedia homeodomain region, and can bind DNA. Furthermore, we observed dose-response cytotoxicity of HepG2 cells with our peptide. No such cytotoxicity was observed in HeLa cells; a morphological change, however, was observed. We examined the cytotoxicity of our drug against 1BR-hTERT normal skin cells. Our peptide drug induced dose-dependent cytotoxicity that was markedly weaker than that of cancer treated cells. Together our data illustrates the isolation of one peptide drug candidate from the AUC Red Sea metagenomic library; furthermore, we were able to observe the selective dose-dependent reduction of HepG2 cell viability

TABLE OF COTENTS

LIST OF TABLES

List of Figures

List of Acronyms Justify the text of the list of abbreviations

ANOVA	Analysis of Variance
APD2	Antimicrobial Peptide Database 2
BLASTp	Basic local alignment search tool-protein
Cgdsr	Cancer Genomics Data Server for r
DNA	Deoxyribonucleic acid
DMEM	Dulbecco's modified essential medium
DMSO	Dimethyl sulfoxate
FBS	Fetal bovine serum
FITC	FLUORESCIN ISOTHIOCYANATE
HMM	Hidden Markov model
MEIS2	Meis1, Myeloid Ecotropic Viral Integration Site 1 Homolog 2
µl	microliter
ml	milliliter
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
PAX6	Paired box protein 6
PBS	Phosphate buffered saline
PI	Promidium iodide
RCF	Relative centrifugal force
RPMI	Roswell Park Memorial Institute
SIX2	Sine Oculis Homeobox Homolog 2
TCGA	The Cancer Genome Atlas

INTRODUCTION

1) Literature Review

Cancer is currently one of the leading causes of mortality and morbidity worldwide. In 2012, about 8 million deaths occurred around the world due to cancer, and about 14 million new cases appeared; the number of new cancer cases per year has been expected to rise by about 70% over the next two decades [1]. Cancer is a term given to a collection of diseases characterized by uncontrolled cell division. The abnormal tumor cells, which arise from normal cells, are able to grow anywhere in the body and can arise from any tissue or cell type. As opposed to benign tumors, malignant tumors (also known as cancers) can spread throughout the body and invade any type of tissue far from the site of the original tumor. The cancer cells use blood vessels and lymph vessels to travel around the body. In contrast to normal cells, cancer cells are undifferentiated; that is, they may not carry out the same specific functions as the original normal tissue cells. Moreover, cancer cells are desensitized to the external chemical signals that regulate their growth, division, and functions. Cancer cells are also able to influence their microenvironment; for example, a cancer mass can influence the development of blood vessels, a process known as angiogenesis, in order to provide itself with nutrients and waste removal. One of the functions of the immune system is removal of the body's aging and/or aberrant cells; cancer cells however, have evolved mechanisms by which they can evade removal by the immune system. Cancer arises through genetic changes in the normal cell (the process of carcinogenesis); those changes can fall within one or a combination of three main classes of genes: proto-oncogenes, tumor suppressors, or DNA repair genes. Proto-oncogenes and tumor suppressor genes are responsible for ensuring normal cell development, growth, and division while DNA repair genes are specialized in repairing any mutations that can arise in the cell's DNA. Mutations to proto-oncogenes and tumor suppressors may occur due to exposure to mutagenic agents such as radiation; failure to repair these mutations by the DNA repair machinery results in the persistence of these mutations and subsequently results in carcinogenesis. As is the case with the other genes, mutations to the DNA repair genes can also develop through environmental stimuli or through aging [2, 3].

The currently used regiments for treating cancer remain problematic. The foremost problem with conventional chemotherapy is the failure to administer the correct amount of the chemotherapeutic agent directly and selectively to the tumor mass or tissue. As a result, the treatment itself may become toxic to the patient. Furthermore, several drug resistance mechanisms have been reported. Once the chemotherapeutic agents have been transported into the cell, intracellular transport mechanisms may re-route the chemotherapeutic agents back out of the cell before they could interact with their intracellular targets. Clearly, there is a need for more selective, more potent, and less toxic cancer therapeutics [4-6].

Towards the latter part of the 20th century, new classes of cancer therapeutics have come into existence (namely, protein-based therapeutics) due to the advancements in biotechnological techniques of that time (such as recombinant protein expression and enhanced protein purification and analysis protocols). These new protein-based therapeutics, or "biologics", offer several advantages over the "small-molecule" drugs: very high selectivity towards their targets, fewer side effects, and subsequently higher potency. Examples of currently used biologics therapeutics include insulin, growth factors, and engineered antibodies [6]. Biologics, however are tedious and expensive to manufacture [6].

Small peptide drugs combine the best of both worlds: they possess the hallmark potency and specificity of biologics, but are much smaller in size, more accessible, and much cheaper to manufacture [6].

Among the classes of peptide drugs, are anticancer peptides [6]. Anticancer peptides are small (5 to 50 amino acids), cationic, amphiphilic peptides [4-6]. The major feature of anticancer peptides that dictates their specificity towards cancer cells is their cationic net charge. In contrast to the normal cells, cancer cells have a relatively negative charge; this relative negative charge is a result of the over-expression of anionic cell surface molecules such as phosphatidylserine, O-glycosylated mucins, siasilated ganglioside, and heparan sulfates [5, 7]. The anticancer peptides are therefore, attracted to the membrane of cancer cells merely

through simple electrostatic interactions. However, membrane fluidity and cholesterol content also affect the accessibility of the peptide into the cell [8].

Two main modes of action have been described for anticancer peptides: membranolytic and non-membranolytic (Figure 1). Anticancer peptides acting through the membranolytic mechanisms disrupt either the cell membrane, mitochondrial membrane, or lysosomal membrane. The anticancer peptides can disrupt the plasma membrane by either the carpet or barrel stave models. In the carpet model, the cationic anticancer peptides align parallel to the cell membrane, and once

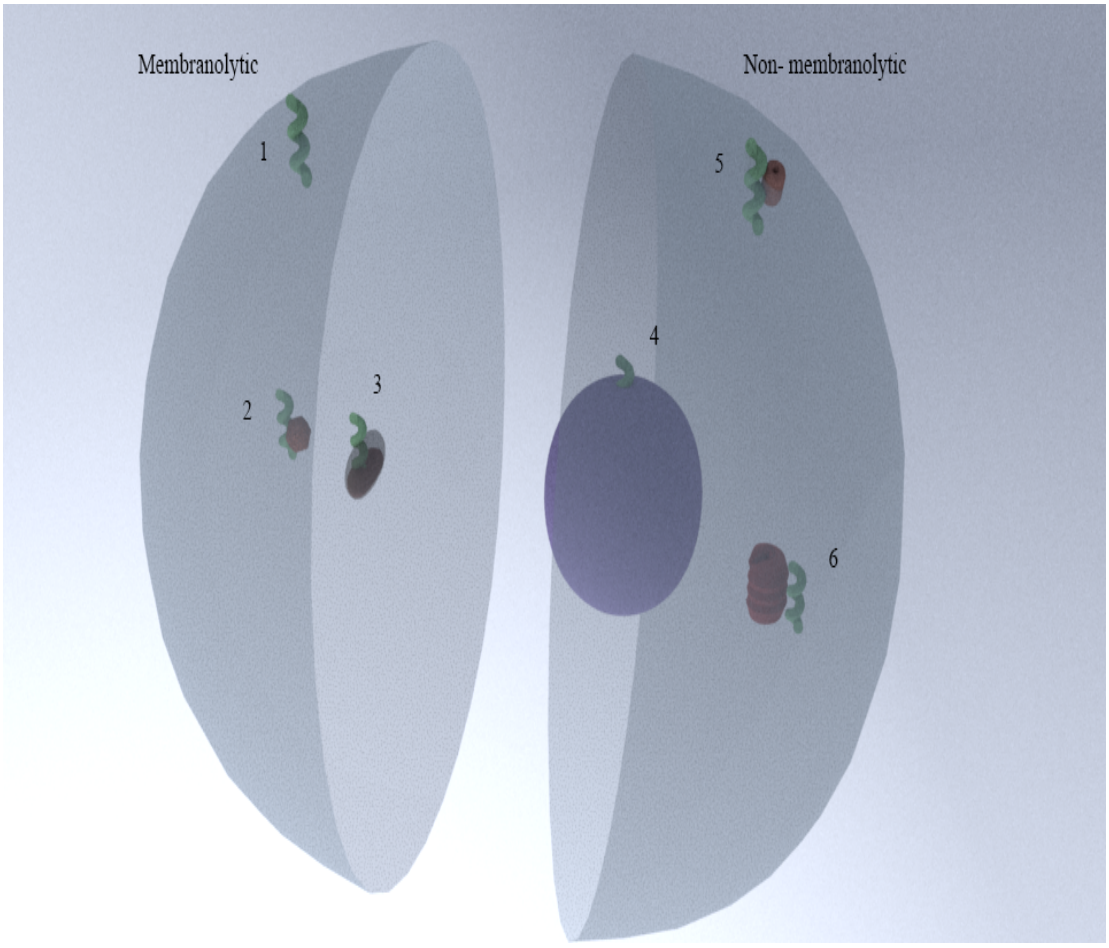


Figure 1: 3D representation of the major modes of action described for known anticancer peptides so far. Mechanisms of action fall broadly into 2 categories: Those which are membranolytic (left half-cell) and those which are non-membranolytic (right half- cell). The membranolytic modes of action affect the membranes of the cell, lysosome, or mitochondria (1, 2, 3, respectively). The non-membranolytic modes of action either modify gene expression of pro-survival genes and induce cell cycle arrest, or affect activity of calcium ion channels and proteasome (4, 5, 6).

a certain threshold concentration of the peptide is reached, the peptides can permeate the plasma membrane. In contrast, anticancer peptides can aggregate, by hydrophobic interactions, to form a structure through the plasma membrane resembling a traditional ion channel. Anticancer peptides can also pass through the plasma membrane and permeate the mitochondrial membrane where they will induce swelling of the mitochondria and release of cytochrome c. Release of cytochrome c can subsequently activate caspase 9 and 3. Furthermore, modification of the lysosomal membrane by anticancer peptides can result in acidification of the cytosol. The non-membranolytic mechanisms of action include activation of calcium channels resulting in calcium ion influx, augmentation of proteasome activity, inhibition of pro-survival genes, or cell cycle arrest [5, 9].

The HPRP-A2 peptide, described by Zhao et. Al. [10], has been found to rapidly induce cytotoxicity in two gastric cancer cell lines by means of cellular membrane destruction (Table 1). They described a combined mechanism by which the HPRP-A2 peptide induces a rise in the intracellular levels of reactive oxygen species and the depolarization of the mitochondrial membrane, which is indicative of mitochondrial damage. The peptide was also able to result in cell cycle arrest. Furthermore, the HPRP-A2 peptide resulted in increased cytotoxicity when used in combination with chemotherapy. In contrast, the Kahalides can trigger modification of the lysosomal membrane which results in cell death [9, 11]. The Kahalides come from an alga, Bryopsis, that the saltwater marine mollusk, *Elysia rufescens* feeds on; Kahalide has undergone phase II clinical trials against solid tumors [12]. Melittin is another anticancer peptide; it has been described to induce apoptosis in HepG2 cells by calcium ion influx through the activation of Ca^{2+}/Calmodulin dependent protein kinase [13]. Other peptides have been described to act in a combination of different modes of action encompassing, in some cases such as Magainin, most of the known mechanisms [9].

Table 1: Mode of action of some representative anti-cancer peptides.

Peptide name	Mechanism of action	Publication
HPRP-A2	Cell membrane desaturation Mitochondrial membrane destruction Cell cycle arrest	Zhao et. Al., 2015
Kahalide	Lysosomal membrane modification	Hamann and Otto, 1996
Melittin	Calcium ion influx	Wang et. Al., 2009
Magainin	Mainly membrane destruction Other modes of action described	Mulder et. Al. 2013

2) Objectives

In this study, we attempt to construct a support vector machine (SVM) pipeline that can detect peptide sequences with potential anticancer activity. We also select the best candidate from the list of potential anticancer peptides and attempt to further examine its potential anticancer activity through protein structure prediction, modelling, visualization, and ligand binding prediction. We Also provide *in vitro* evidence of the anticancer effect of our peptide.

MATERIALS AND METHODS

1) Computational Modeling

A. METAGENOMIC LIBRARY SCREENING AND CANDIDATE PEPTIDE SELECTION

We adapted the method proposed by Tyagi et. Al. 2013 [14] for mining large datasets containing potential anticancer peptide sequences in order to search the AUC Metagenomic library. We compared a dataset of experimentally validated anticancer peptides [15-19] to a dataset of antimicrobial peptides and another dataset of random peptides.

All possible oligopeptide frequencies were investigated in a size range of 1 to 30 amino acids. We also calculated the amino acid and dipeptide (2 amino acids) frequencies for the anticancer and antimicrobial datasets. We compared the frequencies of amino acids and dipeptides in the anticancer and antimicrobial peptide datasets to those of the peptides in the Metagenomic library (t-test, $P<0.05$, Bonferroni multiple testing correction). Only the peptides recognized as anticancer were chosen. A sliding window of increasing size, starting from 5 amino acids, was used to generate the peptides from the Metagenomic library translated reads. The recognized peptides were scored according to how well they conform to the mean amino acid and dipeptide frequencies of the experimentally proven anticancer dataset, that is, they fall within the standard error from the mean.

We searched the passing peptides for presence of Hidden Markov Models (HMM's) [20, 21] previously reported on experimentally verified anticancer peptides.

We confirmed our final predictions using the online tool developed by Tyagi et al [14] (http://crdd.osdd.net/raghava/anticp/).

B. FILTERING A SINGLE CANDIDATE FOR MODELLING

We selected a single anticancer peptide from the final shortlist of potential anticancer peptides for modelling. The criteria used were cationicity, model prediction score, and size.

C. PEPTIDE PERFORMANCE OPTIMIZATION

In order to increase the statistical performance of our peptide, when run against a dataset of experimentally validated and random peptides, we carried out a series of optimization steps. We ran the peptide sequence as a FASTA file on the AntiCP web server for anticancer peptide prediction. We chose model 2 for analysis, which compares the query to a set of experimentally validated anticancer peptides and random peptides, as opposed to a dataset of anticancer and antimicrobial peptides. Amino acid modifications were serially introduced to the peptide in order to maximize the prediction score. We chose only the modifications that occurred outside of the HMM alignment region. We stopped serial modifications as soon as the prediction model returned a score high enough to differentiate the query as an anticancer peptide.

D. BLASTp ALIGNMENTS

Using our peptide as a query against two datasets, we ran a BLASTp search using default parameters, except for the threshold cutoff of 1000, since the sequence is short. One dataset was downloaded from the APD2 web server containing all the experimentally validated antimicrobial/ anticancer peptides [15]; the second BLASTp search was against the NCBI database across *Homo sapiens* using all default parameters in order to test for sequence similarity with human proteins.

E. 3D MODELLING

In order to proceed with the 3D modelling for our peptide, we ran our peptide through I-TASSER, a web server for protein secondary structure prediction [22]. We submitted the query peptide sequence as a FASTA file. Finally, we downloaded I-TASSER output files for modelling and analysis.

We used the modelling and visualization software Chimera, developed by UCSF in order to produce all modelling and visualization figures and numbers [23]. The .pdb files from the protein sequence analysis and for the functional domains of the two most similar proteins predicted by I-TASSER were fed into Chimera and we conducted a series of alignments. We first superimposed all of the sequences using default parameters in the MatchMaker tool within Chimera. We then conducted the structural alignments in Chimera.

F. LIGAND AND BINDING SITE PREDICTION

We conducted ligand and binding site prediction using the web server COACH for ligand prediction [24]. This web server predicts the possible ligands for a query sequence and returns the most likely binding sites and binding site amino acids based on the BioLip database [25]. The .pdb output files were downloaded and we fed them into the Chimera software for protein-ligand visualization and analysis.

G. GENE EXPRESSION ANALYSIS

We used the R statistical programming software in order to investigate the expression patterns of Pax6 and Meis2 (the closest transcription factors to our peptides). We constructed several loops towards this end. We first started by calling on the cgdsr and RCurl libraries within R. We then used the RCurl library to log into The Cancer Genome Atlas (TCGA) link containing only the TCGA study data.

Initially, we looked at Pax6 and Meis2 expression data in general across all TCGA studies. We constructed a loop that obtains the TCGA mRNA sequencing median z-scores. The loop then computes the mean expression scores and deposits the results in a matrix. A box plot was generated from within R to present the results.

Subsequently, we wanted to get a more detailed look at the expression data for Pax6 and Meis2 in each cancer study. We constructed another loop to build a matrix to hold the expression data for the gene of interest in each individual study. We called on second generation RNA sequencing median z-scores in this loop. We Only included the studies containing fully analyzed data. A bar plot was generated from within R to present the data.

Finally, we were interested in studying the differential expression patterns of Pax6 and Meis2 between normal and cancer cells. We obtained level 3 RNA sequencing data (normalized number of reads) for 15 cancer samples with their paired normal tissue from the TCGA data portal. We subsequently constructed a count matrix in R by combining all normal and cancer expression data with their genes into one matrix.

In order to examine differentially expressed genes, we used the EdgeR library. We followed the same workflow pipeline proposed by the Edger User's Guide for comparing cancer versus matched normal RNA-sequencing data.

2) Laboratory Validation

A. DRUG SYNTHESIS AND PREPARATION

The peptide drug was synthesized by GL Biochem LTD, Shanghai, China and shipped as a lyophilized powder. The peptide was synthesized at 98% purity. We stored the lyophilized powder in the -20 freezer in a sealed vial away from light.

We prepared stock solutions of our drug by dissolving the powder in deionized water at a concentration of 1 mg / ml. We stored the stock solutions in 1 ml Eppendorf tubes split into 0.5 ml aliquots.

B. CELL CULTURE

We used HepG2 (high passage number), HeLa cells, and 1BR- hTERT cells in this study. The HepG2 cells were previously purchased from Vacsera, Egypt; they are a hepatocellular carcinoma cell line derived from a hepatocellular carcinoma of a 15-year-old Caucasian male [26]. HeLa cells are a permanent human adenocarcinoma cell line derived from a rare cervical adenocarcinoma of a 30-year-old black female [27, 28].The 1BR-hTERT cell line is an immortalized human skin fibroblast cell line. Both the HeLa and 1BR-hTERT cells were kindly provided by Dr. Andreas Kakarougkas.

We maintained the HepG2 cells in RPMI-1641 media (Lonza) completed with 10% heat inactivated fetal bovine serum (FBS) and 5% Penicillin-Streptomycin. The HeLa cells were kept in DMEM (Lonza) completed with 10% heat inactivated FBS and 5% Penicillin-Streptomycin. The 1BR-hTERT cells were kept in the same conditions as the HeLa cells. We incubated the cell culture flasks at 37 degrees Celsius with 5% CO_2. We observed the cells under the Olympus 1X70 microscope for morphology and cell death.

C. CELL COUNT

In order to count cells for seeding, we used trypan blue staining. 20 µl of cells were mixed thoroughly with 20 µl of trypan blue dye and the cells were counted using a hemocytometer slide over 4 chambers. We calculated the number of cells using the following formula:

$$Number\ of\ cells\ per\ milliliter =$$

$$number\ of\ \ cells\ counted\ x\ (dilution\ factor\ \div$$

$$number\ of\ chambers\ used)\ x\ 10{,}000$$

D. CELL CYTOTOXICITY ASSAY

We used the MTT assay in order to determine cell cytotoxicity after exposure to the test condition. The MTT (3-(4, 5-dimethylthiazolyl-2)-2, 5-diphenyltetrazolium bromide) reagent is a yellow tetrazolium compound which reacts with the mitochondrial dehydrogenase enzymes of viable cells. The reaction of the MTT reagent with the dehydrogenase enzymes forms purple formazan crystals[29]. In order to detect cell viability, as a measure of the cell cytotoxicity, after exposure to the test conditions, we seeded HepG2 or HeLa cells in a 96- well plate (Corning, USA) at a density of 1 x 10^4 cells in 100 µl of fresh complete media. After 24 hours from seeding, we discarded the old media and added the peptide drug at concentrations of 512, 256, 128, 64, 32, and 16 µg / ml diluted in fresh media by serial dilutions. 24 hours following addition of the drug, we replaced the old media with fresh media containing 0.5 mg/ml MTT (5mg/ml stock) (Serva, Germany) to each well. We incubated the plate for 3 hours. We then discarded all the media and added 100 µl of DMSO (Sigma-Aldrich, USA) to each well in order to dissolve the formazan crystals. We incubated the plate at room temperature, wrapped in Aluminum foil for 15 minutes. We measured absorbance at 490 nm using the BMG Labtech Spectrostar Nano plate reader.

We calculated cell viability by subtracting the blank absorbance reading from all other wells. We then divided the sample absorbance reading by the control absorbance reading then multiplied by 100.

For all experiments, we used several control conditions as follows: blank (media with no cells), cells with complete media only, and cells with media and the solvent used to prepare the drugs employed in the study. The solvent control was prepared only for the highest drug concentration (i.e. as if preparing the 512 μg/ml but with no drug). We subsequently converted mass volumes to molar volumes.

E. CELL DEATH ASSAY

We used the AlexaFlour 488 AnnexinV/ Dead Cell kit from ThermoFischer Scientific in order to evaluate the mode of cell death in the treated cells. During the highly regulated and programmed mode of cell death (apoptosis), the apoptotic cells present phosphatidyl serine (PS) on the outer leaflet of their plasma membranes. Annexin V, a human anticoagulant and phospholipid binding protein can readily bind PS presented on the outside of the apoptotic cells. When conjugated with a fluorophore, Annexin V can emit green light when excited with ultraviolet light. Moreover, propidium iodide (PI) is a fluorophore that can tightly bind nucleic acids. Live and apoptotic cells are impermeable to PI. PI, however, can permeate into dead cells and bind their nucleic acids. When excited with ultraviolet light, PI emits red light [30, 31].

We plated $0.3x10^6$ HepG2 cells per well in a 6- well plate. We plated wells for untreated control and drug treated (15.2 μM) conditions in duplicate. After 24 hours we discarded the spent media and added the peptide drug to the treatment wells. After another 24 hours we collected the media containing the floating cells and deposited them in a 15ml falcon tube. We subsequently trypsinised the adherent cells and combined those cells with the dead cells in the same falcon tube. We then pelleted the cells by centrifugation for 5 minutes at 500 RCF 4 degrees Celsius. The supernatant was discarded and the cells were re-suspended in 1 ml fresh media. 20 μl of the cell suspension was used for counting. We then collected the appropriate amount of cell suspension in order to obtain 50,000 cells. We pelleted this suspension and washed with 100 μl PBS. Again, we pelleted the cells and re-suspended in 1x Annexin binding buffer. We added 2 μl of the Annexin V reagent to the cell suspension in addition to 1 μl of PI. We incubated the cells suspension in the dark for

15 minutes at room temperature. 25 µl of the resulting cell suspension was then transferred to a glass slide and a cover slip was carefully placed over the droplet. We visualized the cells under the microscope using the appropriate FITC filters.

3) Statistical Analysis

In order to calculate mean viability readings and standard deviation, we used the raw un-blanked absorbance readings as input to the R statistical analysis software. We trimmed the unused wells out of the absorbance readings table and blanked all the absorbance readings. We also calculated the mean viability for each treatment.

We used the GraphPad Prism software to calculate the half maximal drug response values (EC_{50}) and to draw the dose response curves. We entered the raw viability data for each set of replicates in each experimental condition along with the different concentrations used. After the drug concentrations were log transformed and the data normalized, we calculated the EC_{50} values. Furthermore, we tested the data for normality and subsequently carried out either parametric or non-parametric two- way ANOVA with multiple comparisons in order to obtain significance values for the viabilities across the different drug concentrations.

RESULTS

1) A predicted 37-residue anticancer peptide

We screened the AUC Metagenomic library across translated sequences for potential anticancer peptides. We obtained a list of 59 potential anticancer peptides, in terms of anticancer versus antimicrobial and anticancer versus random peptide scoring performance. While applying the filtering methods described, any anionic peptides were disregarded and any peptides with a size larger than 50 amino acids were also disregarded (since most anticancer peptides are cationic and fall within the size range between 5 – 50 amino acids, roughly [5, 6]); two candidate peptides remained. Both peptides contained HMM alignment profiles indicating presence of a homeodomain (PF00046.). One peptide was 30 residues long while the other was 37. We chose the 37 residue peptide, even though the shorter peptide was more cationic since the homeodomain HMM occurred on residue 0 to 30 on the shorter peptide; this would present a problem when optimizing the peptide since any mutations introduced would likely influence the HMM alignment. This 37-mer peptide came from the Atlantis II Deep brine pool sub-seafloor sediment section 6; the original sequence of which can be found on the NCBI Short Sequence Archive (SRA) under the name of the American University in Cairo along with the entire American University in Cairo Red Sea Metagenomic Library. The original peptide sequence was TKEQKEQIAKATGLTTKQVRNWYVQLNASIKVMLTSI (Table 2). The modified and optimized peptide sequence contains 3 Cysteins at positions 33, 34, and 36 in place of the original amino acids. The optimization of the original peptide resulted in an increase in the model performance score; furthermore, the hydropathicity of the peptide increased (Table 3).

Table 2: Anticacancer Peptide Drug Information.

Nature of Data	Data
Dataset	ATII-6Sediments_qc-prot.acp
SVM Score	3
Peptide length	37
HMM id	Homeobox
HMM accession	PF00046.24
HMM start	1
HMM end	31
ACP/AMP model prediction (Tyagi, Atul, et al. 2013)	0.74
ACP/NON-ACP model prediction (Tyagi, Atul, et al. 2013)	-0.98
Hydrophobicity	-0.18
Hydropathicity	-0.39
Hydrophilicity	0
Charge	4
Molecular weight	4334.71

Table 3: The new chemical properties of the modified anticancer peptide

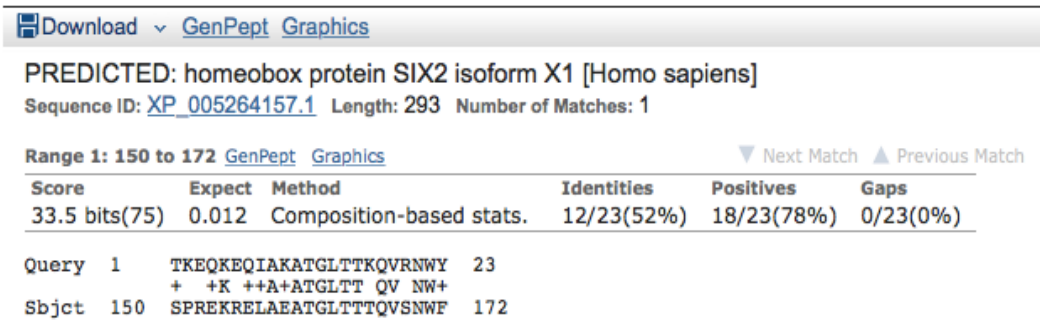
	ACP/NON-ACP model prediction (Tyagi, Atul, et al. 2013)	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge
Original	-0.98	-0.18	-0.39	0	4
Modified	0.28	-0.2	-0.34	-0.01	4

2) The predicted anticancer peptide is a homeodomain protein that aligns with an Arthropod defensin

We ran the chosen peptide sequence through a BLASTp search against two datasets; one contained a list of experimentally validated anticancer/random peptides from the AntiCP web server; the second dataset contained a list of experimentally validated anticancer peptides from the APD2 peptide library. For the AntiCP dataset,

our peptide aligned best with an anticancer/ anti-Gram+ peptide from *Drosophila virilis* [32] (coverage= 24%, identity = 56%, E-value = 7.1) . The results for the second alignment yielded an alignment with an Arthropod defensin from Stomoxys calcitrans (the stable fly) [33] (coverage = 72%, identity = 33%, E-value = 9). We submitted this peptide to the AntiCP server anticancer peptide prediction tool and it predicted that this peptide is also an anticancer peptide; We also ran this sequence as a query on the PFAM web tool and it confirmed that this sequence is that of an arthropod defensin.

We ran our peptide through a third BLASTp search. This instance, however, we ran our peptide against the NCBI non-redundant (nr) database and included only *Homo sapiens* in the search. The best alignment was with the human homeobox gene Six2 (E-value = 0.012) (Figure 2).


Download GenPept Graphics

PREDICTED: homeobox protein SIX2 isoform X1 [Homo sapiens]

Sequence ID: XP_005264157.1 Length: 293 Number of Matches: 1

Range 1: 150 to 172 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
33.5 bits(75)	0.012	Composition-based stats.	12/23(52%)	18/23(78%)	0/23(0%)

```
Query  1    TKEQKEQIAKATGLTTKQVRNWY  23
            +  +K ++A+ATGLTT QV NW+
Sbjct  150  SPREKRELAEATGLTTTQVSNWF  172
```

Figure 2: Summary of the best alignment returned by BLASTp against our peptide.

3) The peptide contains a helix turn helix structure and is structurally similar to Pax6

In preparation for modeling and visualization of the predicted anticancer peptide, we submitted the peptide sequence to I-TASSER, a web server for protein secondary structure prediction. I-TASSER results indicated that our peptide consisted, structurally, of two helices separated by a coiled region (C-score = -0.17); the two helices lay roughly perpendicular (about 79 degrees) to one another, in accordance to the consensus geometry of the homeodomain[34] (Figure 3). The protein

structurally closest to our peptide is the homeodomain of the human paired box protein, Pax6 (TM-score = 0.815, RMSD = 1.39, coverage = 1.000) (Figure 4).

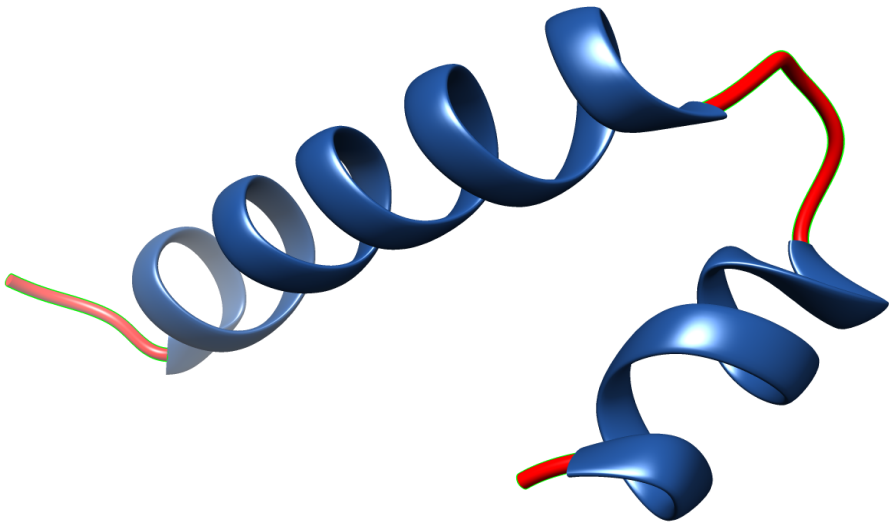


Figure 3: 3D rendering of our peptide drug with two alpha helices forming an angle of about 79 degrees.

Proteins structurally close to the target in the PDB (as identified by TM-align)

Top 10 Identified stuctural analogs in PDB

Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
	1	4cxfA2	0.820	1.16	0.189	0.919	Download
	2	2cueA	0.815	1.39	0.135	1.000	Download
	3	3gg7A	0.812	0.56	0.121	0.892	Download
	4	1ub2A	0.812	1.13	0.216	1.000	Download
	5	1fjlA	0.796	1.35	0.216	1.000	Download
	6	1x2nA	0.796	1.24	0.351	1.000	Download
	7	4qpoA	0.794	0.93	0.086	0.946	Download
	8	3hugG	0.791	1.55	0.162	0.946	Download
	9	3veaB2	0.788	1.20	0.081	1.000	Download
	10	3nauA	0.787	1.39	0.135	1.000	Download

Pax6

Figure 4: Top 10 structurally closest proteins to our peptide as predicted by TM-align tool within I-TASSER. We choose the second best structure (structure arranged in descending order) since the first structure does not cover the entire length of the sequence.

As a confirmatory step, we obtained the homeodomain structure of Meis2 (one of the proteins used by I-TASSER as a template to build our peptide structure, also a human homeodomain protein) (Figure 5) protein and Pax6 and subsequently fed them into Chimera with our peptide and we carried out two alignments: our peptide against Pax6 and our peptide against Meis2. Our peptide aligned, in terms of sequence (within 5 angstroms across all residues), with Meis2 at 42.42% identity across non-gap stretches. On the other hand, Pax6 aligned with our peptide at only 13.5% identity across non-gap stretches.

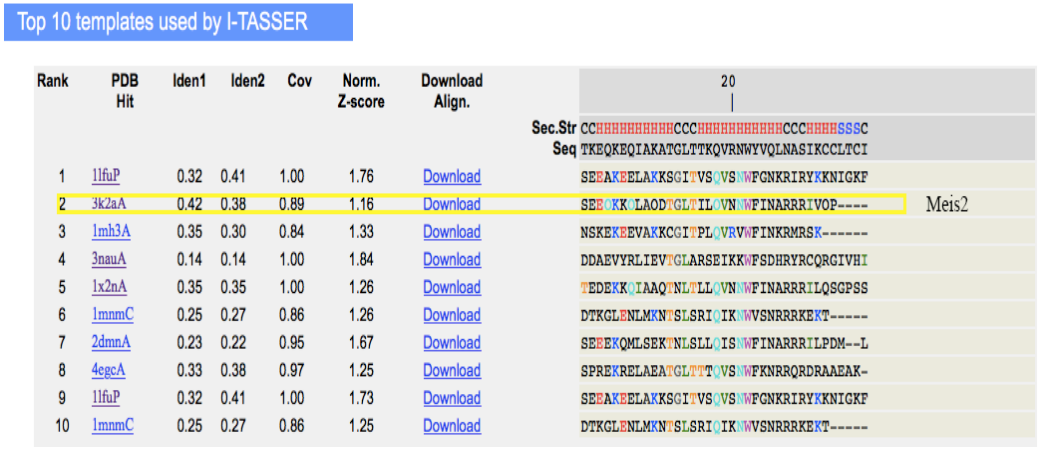


Figure 5: Top 10 templates used by I-TASSER to build the structure of our peptide. We chose the second best template (Meis2 – results are arranged in descending rank order) since it is a human protein, as opposed to PBX (the best alignment) since it is a mouse protein.

The Pax6 homeodomain structure, however, was the closest to our peptide with a plane angle between the planes of the C-terminal recognition helix and the two short N-terminal helices of 78.3 degrees (figure 6b), as opposed to 79 degrees in our peptide (figure 6a); Meis2 plane angles, however, were less congruent with our peptide, at 62.3 degrees (Figure 6c).

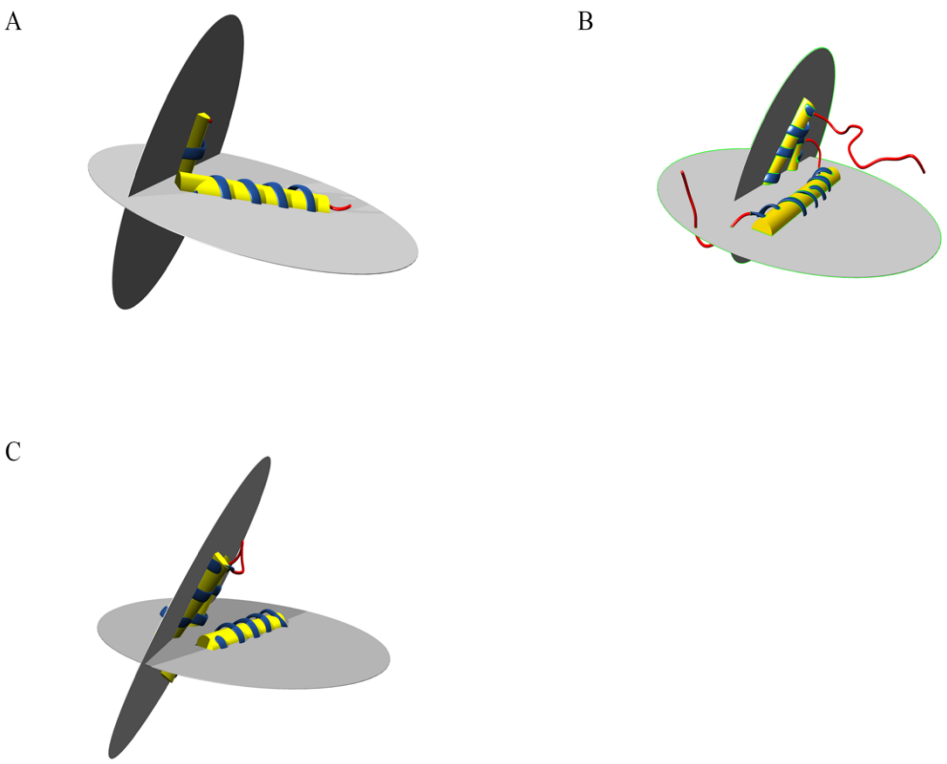


Figure 6: Plane angles formed between the recognition helices and N-terminal helices of: A- our peptide, B- Pax6, and C- Meis2. The planes themselves simply indicate the plane formed by each helix for ease of viewing.

4) Our peptide binds DNA over the same sequences as the antennapedia homeodomain

We wanted to find out the potential ligands for our peptide; therefore, we submitted the structure file of our peptide (output by I-TASSER), to the ligand prediction software, COACH. The most significant result was the output from the COFACTOR tool within COACH. Our peptide was shown to bind nucleic acids, with residues 5, 20, 24, 27, 28, 31 forming the binding site (C-score = 0.1). The closest protein-nucleic acid set used by COACH prediction as a template for ligand predication was from the Antennapedia homeodomain structure bound to DNA (PDB 9antA, TM-score = 0.767, RMSD = 0.81, coverage = 0.919). Using this model, we were able to infer the DNA sequence to which our peptide would most likely interact: 5'-AGAAAGCCATTAGAG-3' (Figure 7). This sequence contains the homeodomain- specific recognition sequence 5'-ATTA-3'.

Figure 7: 3D representation of our peptide (turqoise) aligned with the Antennapedia homeodomain. The C-terminal recognition helix is aligned with the major groove of the DNA molecule in the same orientation as its template indicating sequence- specific interaction potential.

5) Pax6 and Meis2 expression patterns across several cancers over TCGA studies

Upon analyzing the expression data for Pax6 and Meis2, which we compiled from TCGA, we observed that Pax6, on average, displays weak underexpression patterns in the TCGA studies. We also observed that Meis2 shows a much more pronounced average underexpression across the TCGA studies, on average (Figure 8).

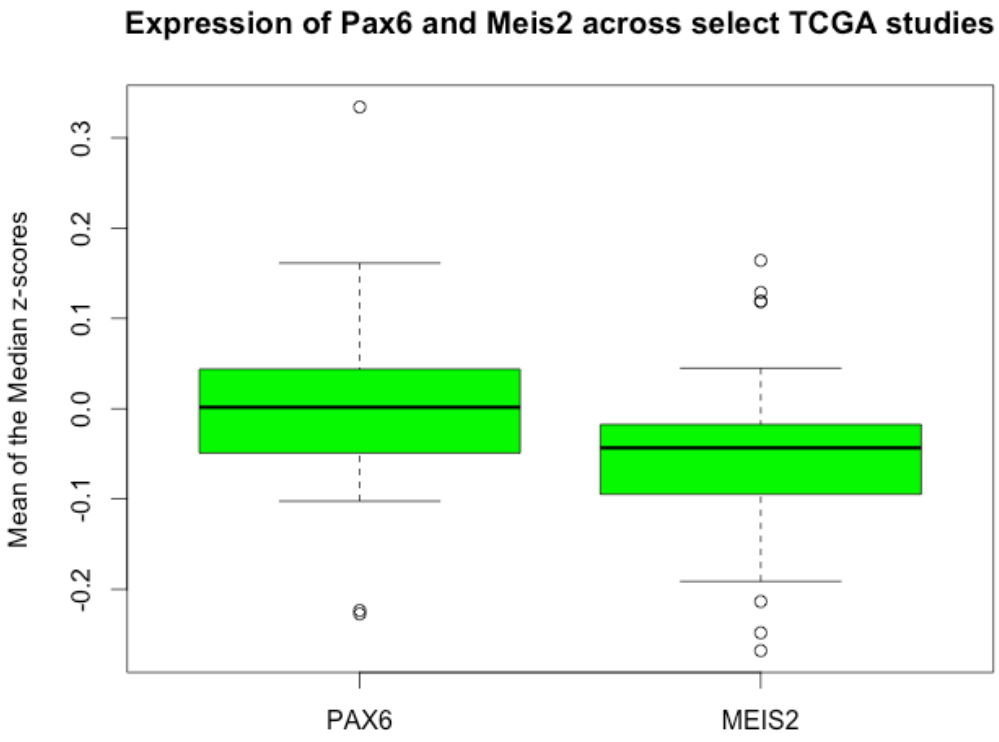


Figure 8: Average gene expression patterns for Pax6 and Meis2.

When taking a closer look (Figure 9), we observed that Pax6 was overexpressed in breast and liver tissue, while being underexpressed in cervical tissue. Meis2, in contrast, was downregulated in breast and cervical tissue, while being upregulated in liver tissue.

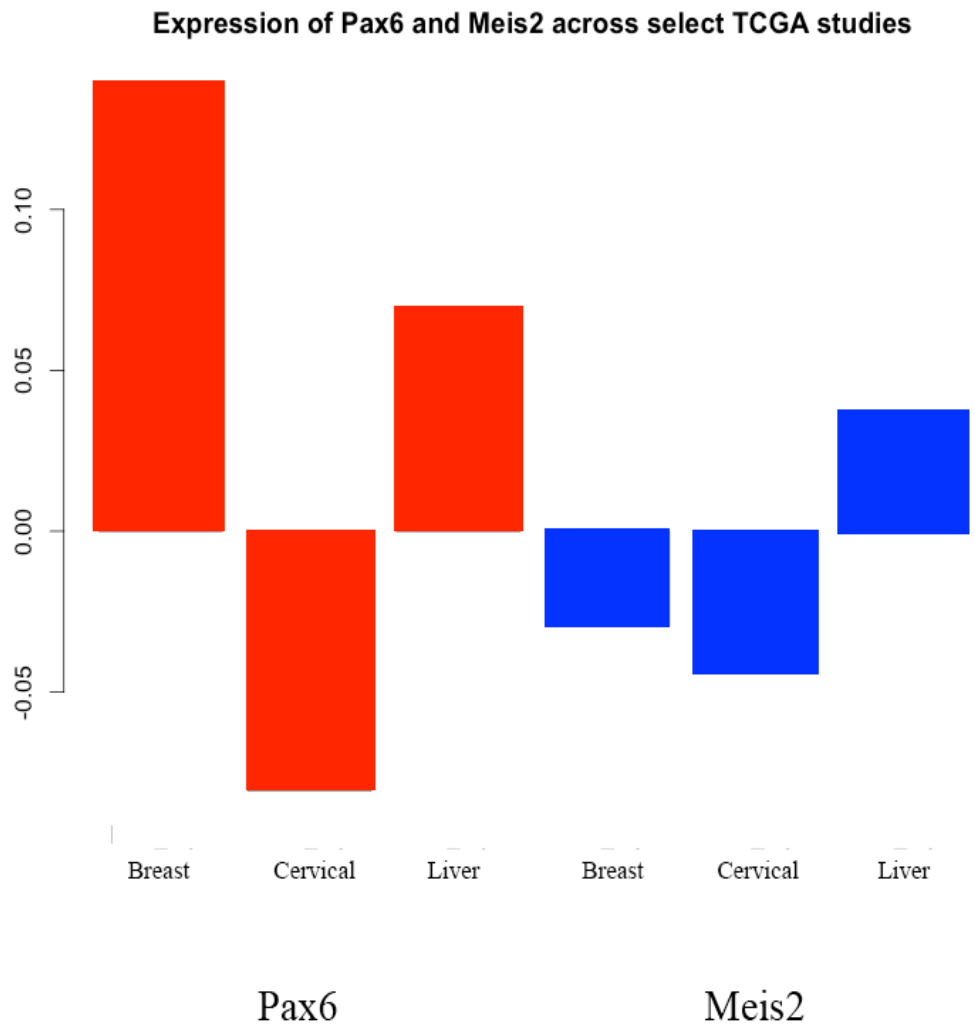


Figure 9: Expression levels for Pax6 and Meis2 in select cancer types representing the cell lines most widely used in our laboratory.

We were also interested in confirming the differential expression patterns of Pax6 and Meis2 in Liver Hepatocellular Carcinoma, represented by the HepG2 cell line in our laboratory.

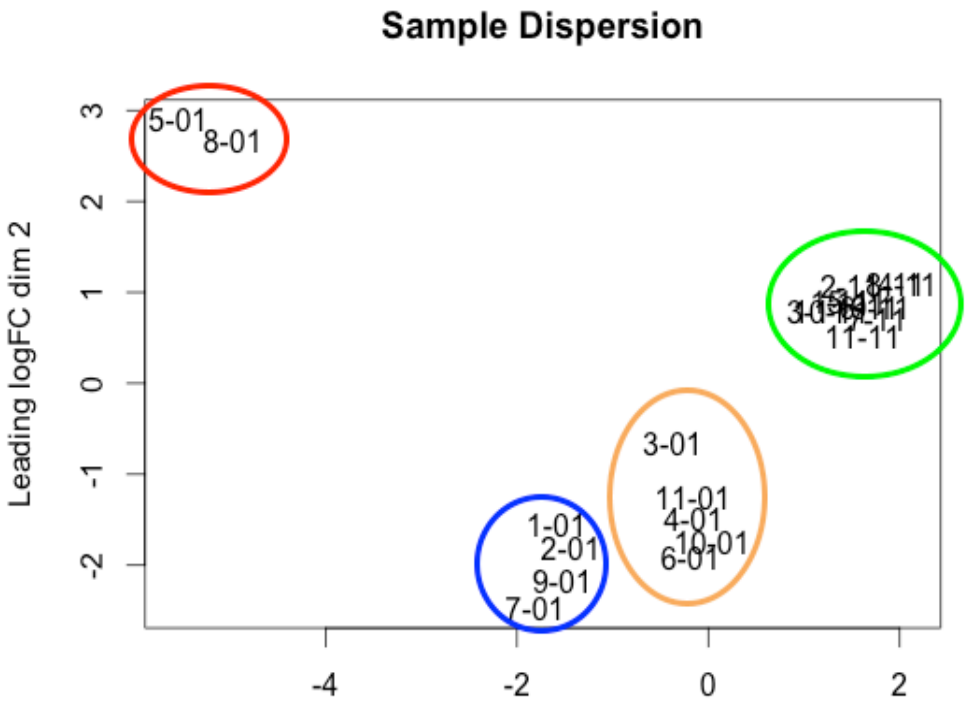


Figure 10: Sample dispersion according to patterns of gene expression for all genes in each HCC sample. The axes represent the leading log of fold change. The samples for the cancer tissue red, blue, and orange circles) appeared more dispersed and heterogeneous indicating varying expression levels. The paired normal tissue samples (green circle) were tightly grouped and more homogenous, indicating consistent expression patterns.

We examined the data obtained and observed, as expected, that the normal samples were more homogeneous and tightly grouped than the tumor samples (Figure 10). The cancer samples were clustered chiefly into 3 major groups; we tried searching in the patient data for a link between the clusters, however, we were not able to find a clear link (Table 4).

Table 4: Hepatocellular Carcinoma Patient Information

Cluster	Sample Nr.	Case	Gender	Age	Stage
Blue	1	TCGA-BD-A3EP	Female	75	I
Blue	2	TCGA-DD-A3A3	Male	45	I
Orange	3	TCGA-DD-A3A4	Male	37	IIIa
Orange	4	TCGA-DD-A3A5	Female	66	III
Red	5	TCGA-DD-A3A6	Female	72	II
Orange	6	TCGA-DD-A3A8	Male	75	II
Blue	7	TCGA-EP-A3RK	Male	73	IIIa
Red	8	TCGA-FV-A3I0	Female	76	II
Blue	9	TCGA-FV-A3I1	Female	NA	II
Orange	10	TCGA-FV-A3R2	Male	75	I
Orange	11	TCGA-G3-A3CH	Male	53	IIIa

We observed that Pax6 and Meis2 were, indeed, expressed in liver hepatocellular carcinoma; however, not differentially expressed ($P<0.05$) between normal and tumor samples (Table 5).

Table 5: Differential Gene Expression Data for Pax6 and Meis2.

Genes	log Fold Change	PValue
PAX6	0.158089913	0.716262029
MEIS2	0.742735053	0.171200501

6) Dose- dependent cytotoxicity upon treatment of HepG2 cells.

We treated the HepG2 cells with our peptide at concentrations ranging from 512 to 3.7 μM. We noticed dose- dependent cytotoxicity with increasing drug dosage (Figure 11). At the highest concentration, 121.5 μM, we obtained around 34% viability. The viability gradually increased to about 73% for the lowest drug

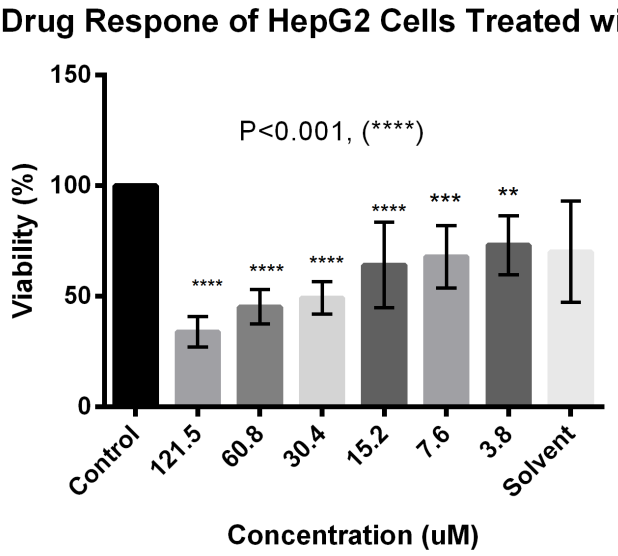


Figure 11:Cytotoxicity drug response curve of HepG2 cells treated with our peptide drug. Significance values are indicated relative to the control.

concentration of 3.8 μM. The concentrations which displayed statistically significant results were 121.5 – 3.8 μM ($P < 0.001$). The EC_{50} for our peptide drug was calculated to be about 8.6 μM.

Upon examination of the drug treated HepG2 cells under the microscope, we observed a vast morphological difference between control cells, and cells treated with our peptide (Figure 12). Peptide-treated cells, however, looked enlarged, the cell

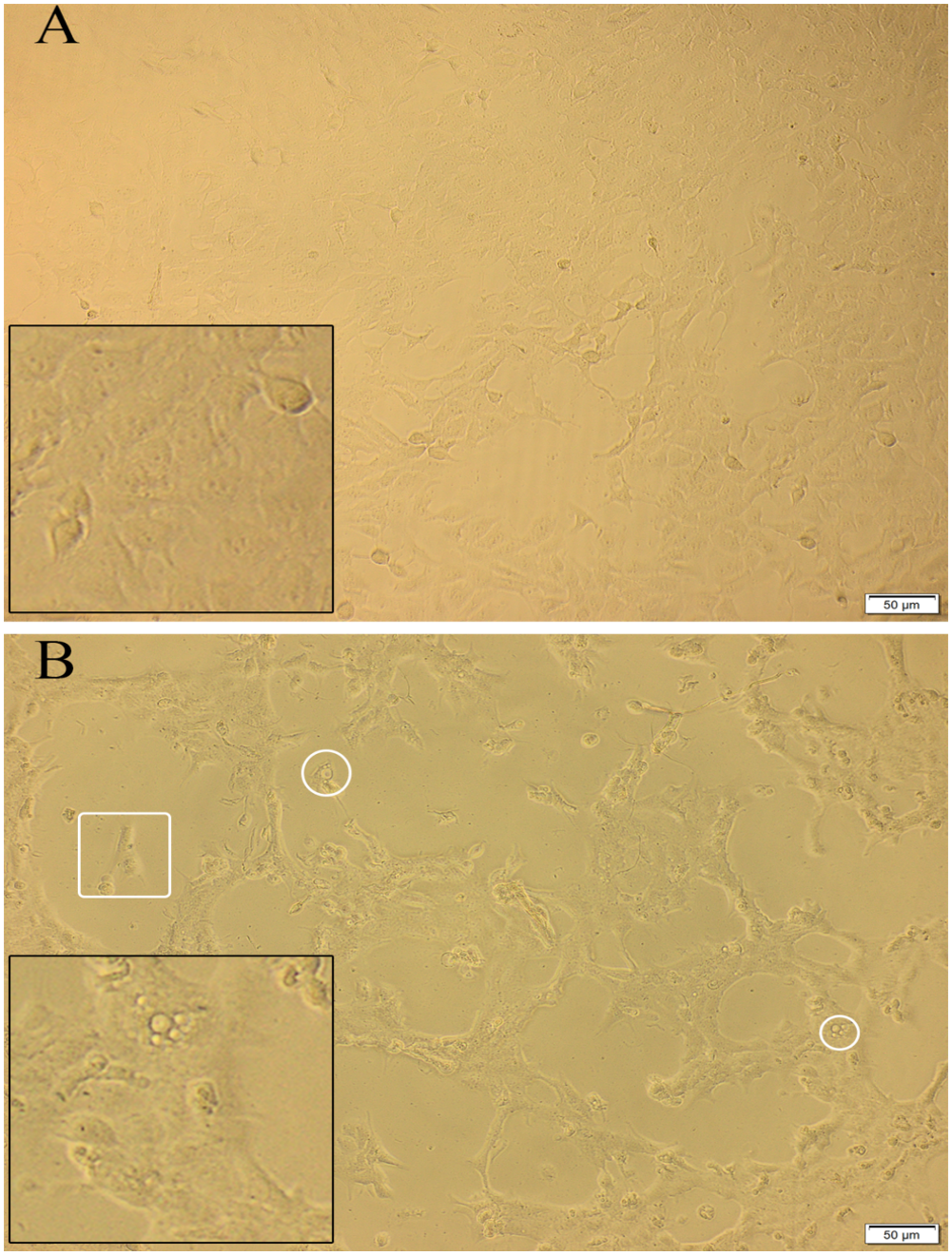


Figure 12: Morphological changes in HepG2 cells induced by our peptide drug. A- Control cells incubated with media only; B- Cells treated with 121.5 µM of our peptide drug. The peptide drug treated cells, at the highest concentration, appeared enlarged and had irregular shapes (rectangle) in addition to a few cells with vacuoles indicated by the circles.

surface looked rough, and the cells displayed multiple extensions. We also observed a few vacuoles in the cells of the highest concentration. The morphological changes gradually subsided with decreasing drug concentration.

7) Irregular cytotoxicity response of HeLa cells to treatment with the peptide drug

In contrast to HepG2, Treatment of HeLa cells with our peptide (Figure 13) drug did not elicit the same pronounced dose- response. At the highest concentration (121.5 μM), our drug resulted in 73.5% viability; the lowest concentration (3.8 μM) resulted in 87% viability. None of the treatment concentrations yielded statistically significant differences in viability with a P- value below 0.001.

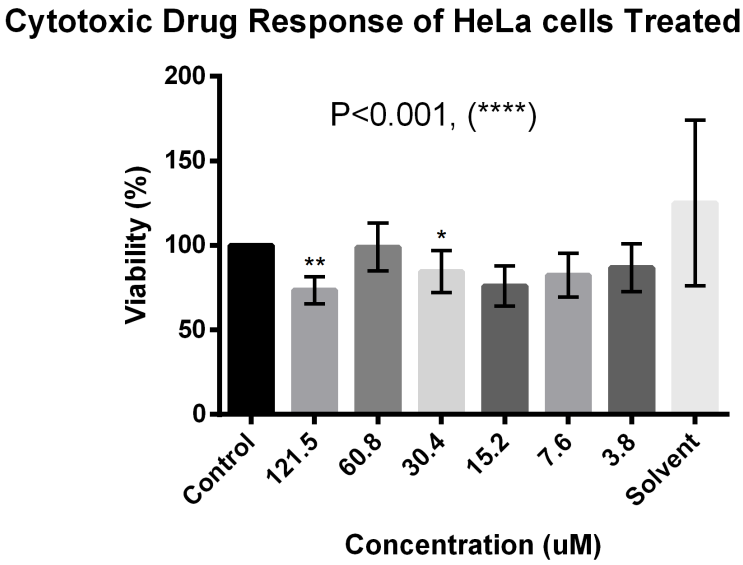


Figure 13: Cytotoxicity drug-response curve of HeLa cells treated with our peptide drug.

Upon examination of the drug treated cells under the microscope, we were able to observe many vacuoles in the cells treated with the highest concentration (Figure 14).

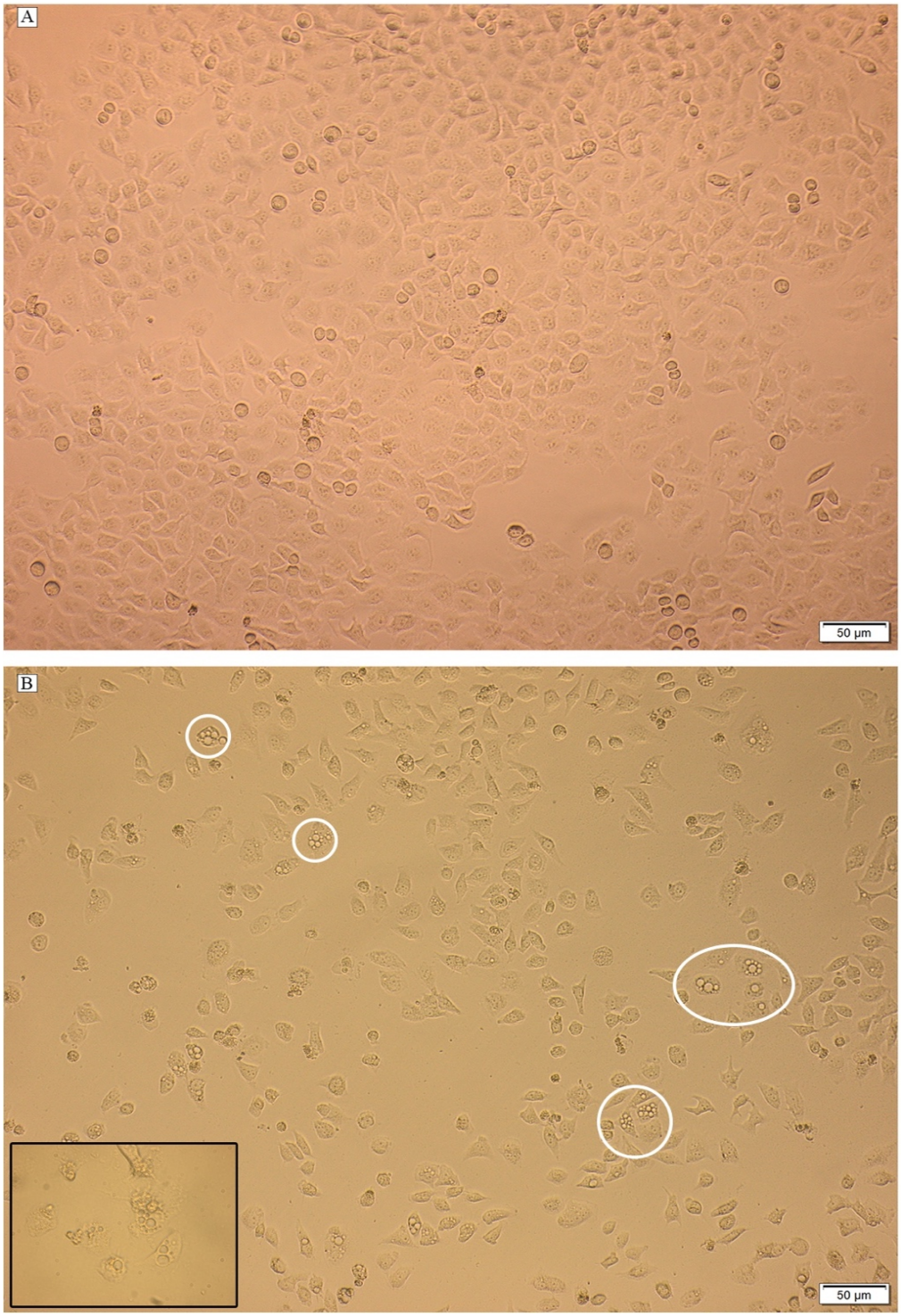


Figure 14: Microscopy for HeLa cells treated with our peptide drug. A- control cells only with growth media; B- Cells treated with our peptide drug at 512 ug/ml. The treated cells were relatively morphologically different and displayed many vacuoles (red circles).

Some cells displayed irregular morphology as compared to the untreated control.

8) Cytotoxicity response of the 1BR-hTERT cells incubated with the peptide drug

We observed a dose- dependent reduction in the viability of the 1BR-hTERT cells upon treatment with increasing concentration of our drug (Figure 15). The highest concentration yielded a viability of about 52%; the viability gradually increased to approximately 88% at 15.2 μM, where the viability plateaued for the remaining concentrations. The EC_{50} for the 1BR-hTERT cells was calculated to be approximately 20.1 μM. The concentration of the drug used on the 1BR-hTERT cells that was closest to the EC_{50} on HepG2 (7.6 μM) resulted in a non-significant drop in viability as compared to the untreated control (that is, our peptide drug is not toxic to normal cells at its HepG2 $EC_{50)}$).

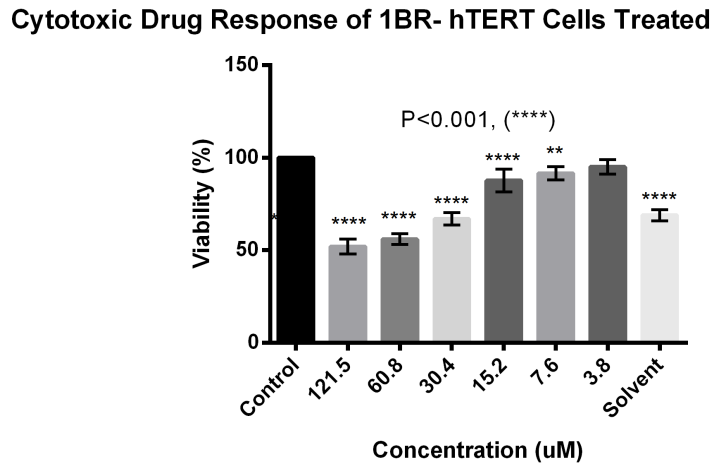


Figure 15: Cytotoxicity drug-response curve for 1BR-hTERT cells treated with our peptide drug. All significance values were calculated relative to the control. only the concentrations from 121.5 to 15.2 μM displayed significant differences

Upon examination of the cells under the microscope, we were able to observe a slight morphological difference between the treated and untreated cells (Figure 16). The treated cells at the highest concentration appeared irregular and shrunken. We

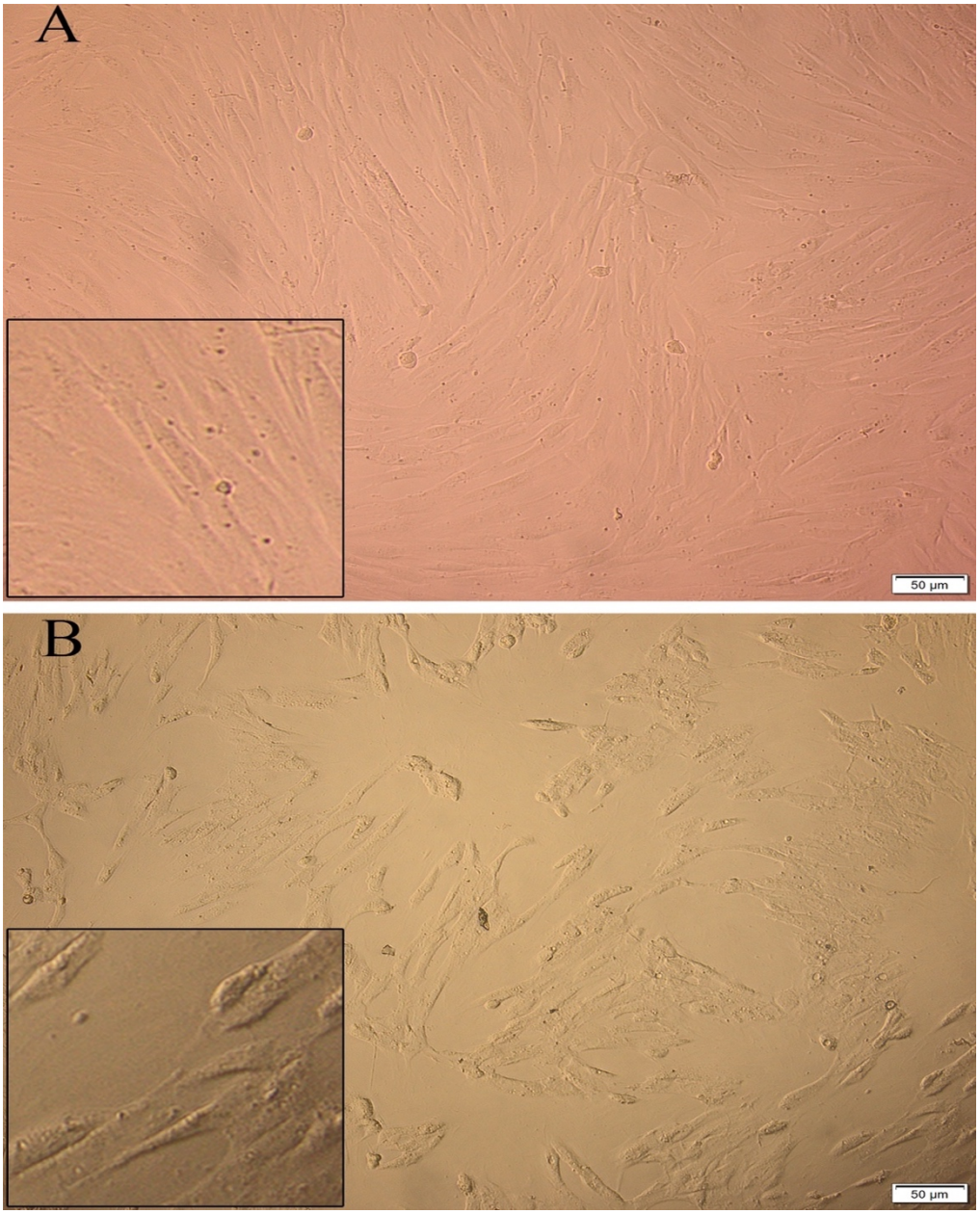


Figure 16: Microscopy images of 1BR-hTERT cells treated with our peptide drug. A- control cells with growth media only. B- cells treated with 121.5 µM of drug. The cells treated with the highest concentration appeared to be smaller and irregular in comparison to the control; those effects gradually subsided with decreasing concentration.

observed that this morphology subsided gradually down to the 15.2 µM concentration where the cells were only slightly smaller than the untreated cells.

9) Increase in apoptotic activity upon treatment of HepG2 cells with the peptide drug

We qualitatively examined the mode of death for the HepG2 cells treated with the peptide drug using ThermoFischer's AlexaFlour488 Annexin V/ Dead Cell assay. We observed a marked increase in apoptotic activity between treated and untreated cells (Figure 17). Furthermore, we observed that the apoptotic cells from the drug treated sample could be divided into two major groups: early and late apoptotic cells. The early apoptotic cells appeared as bright green with faint or no red signal from PI staining. The majority of the cells, however, were late apoptotic; these cells presented as green-stained rings with a strong red signal in the center. We counted 1 early apoptotic cell in the control from about 45 cells in bright field. In contrast, we counted 12 late apoptotic cells and 15 early apoptotic cells (with a total of 27 cells in bright field) for the drug treated condition, in the microscope viewing field.

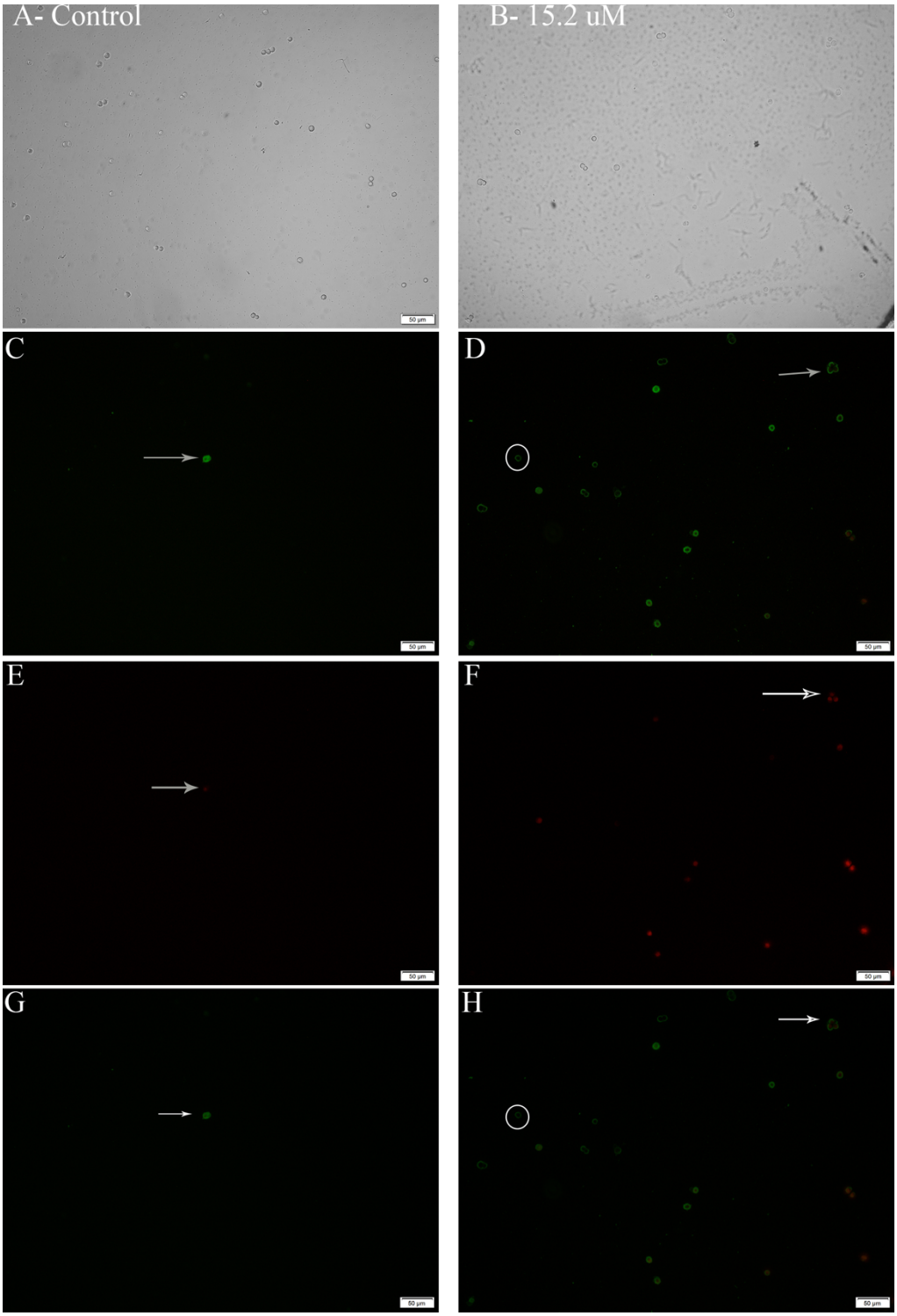


Figure 17: Annexin V/ PI staining of HepG2 cells treated with our anticancer peptide drug. Left column is the control, Right column is the drug treated cells. A, B- bright field; C, D- Annexin staining (the arrows point to cells of late apoptosis while the circles indicate early apoptotic cells; E, F- PI staining; G, H- combined channels. These images indicate an increase in apoptotic activity in the drug treated cells as opposed to the control.

DISCUSSION

Peptide based therapeutics are currently gaining attention as a replacement, or at least, a complement to small molecule therapeutics [4, 14, 35]. Large peptide drugs, or biologics, circumvent the off target toxicity problem of small molecule drugs and also offer much stronger action; that is, since biologics are more specific and on-target, a much lower dose is needed to achieve the same effect. Biologics, however, are expensive and tedious to produce. Small peptide drugs fall within the size gap between small molecule drugs and biologics (0.5 to 5 kDa); they combine the small size of small molecule drugs, and also the potency and specificity of biologics, while also being relatively inexpensive and relatively less tedious to produce[4-6].

Small peptides with anticancer activity are cationic, amphiphilic, peptides containing 5 to 50 amino acid residues. The cationic nature of those peptides confer their specificity towards the relatively anionic cancer cells, in contrast to small molecule chemotherapeutics while amphiphilicity aids in the plasma membrane permeation [5].

We were able to develop a workflow for detecting potential anticancer peptide sequences from the AUC Red Sea Metagenomic library using several support vector machines (SVM's). Our SVM pipeline was able to identify a list of possible anticancer peptides from the AUC database. We filtered the resultant list of potential anticancer peptides for the most promising candidate. Furthermore, we optimized our peptide sequence in order to improve its performance in the SVM model. We believe that none of the peptides were initially recognized as anticancer in our model because more variation between anticancer and random peptide sequences was needed for the model to recognize the peptide as being anticancer (i.e. a larger sample size with more diverse sequences). The first SVM model (anticancer versus antimicrobial) was able to identify the anticancer peptides right away since we were comparing two different classes of small peptides. That is, comparing two different classes of proteins is easier in terms of SVM models as opposed to comparing a peptide to all other peptides.

We subsequently set out to gather some information about our peptide. We conducted a few BLASTp alignments on NCBI. The most significant searches suggested that the peptide aligns with an arthropod defensin and a human homeodomain protein, Meis2. These results suggest that our peptide may also possess some antimicrobial properties. Furthermore, the BLASTp results indicated that the peptide might potentially have transcriptional regulatory activity within the tumor cell; these results were confirmed by the HMM alignment which also indicated that the peptide is, in fact, a DNA binding homeodomain peptide.

Secondary peptide structure prediction using I-TASSER combined with visualization and modelling using Chimera confirmed the presence of the DNA-binding region of the homeodomain structure. The peptide contains two alpha helices separated by a coiled region. The size-asymmetric helices lay at an almost perpendicular angle to one another; a feature typical to homeodomain proteins[34]. Our peptide, however, lacks a third N-terminal short helix; put together, we concluded that our peptide contains only the C-terminal DNA binding portion of the homeodomain.

Ligand prediction and modelling using COACH and Chimera confirmed the DNA binding activity of our peptide. We observed that the longer C-terminal helix does the actual peptide-DNA interaction. The closest template model used for ligand prediction by COACH was that of an Antennapedia homeodomain bound to DNA. When we observed that our peptide might possibly have a transcriptional regulatory mode of action we were concerned that it may not be able to localize inside the nucleus where it exerts its effect. The observation, however, that our peptide aligns with the antennapedia homeodomain sequence was reassuring since some research indicates that portion of the antennapedia homeodomain can be used as a nuclear localization signal in order to direct cell-penetrating peptides into the nucleus[36, 37]. In addition, we observed that our peptide is able to bind to DNA at the same sequence as the Antennapedia homeodomain.

Based on the computational results, we initially hypothesized that our peptide will probably either competitively bind to the Pax6 or Meis2 target promoter sequences

and inhibit their effect or it will replace the action of inactive or mutated Pax6 or Meis2. However, since our peptide was able to elicit dose-dependent cytotoxicity in HepG2 cells and induce morphological changes in HeLa cells (two cancers with opposing expression patterns of Pax6 and Meis2), we refuted this scenario. We currently hypothesize that our peptide drug may be acting through one of the membranolytic modes of action; namely, the mitochondrial mediated apoptosis pathway. In order to confirm this hypothesis, it would be advisable to investigate the levels of active caspase 9 and 3 and oligomerized Apaf-1 since they are indicative of the mitochondrial mediated apoptotic pathway [5].

We also observed that the solvent in the 1BR-hTERT experiment did result in a significant drop in cell viability ($P<0.001$). We attribute this discrepancy to their being less growth medium available for the cells at the highest concentration (since the solvent control was prepared at the highest concentration) where the solution was about 50% deionized water and 50% media. This effect subsided once the drug concentration (and subsequently the amount of deionized water) decreased. There was, however, some degree of cell death that can be attributed to the drug since there was a significant difference ($P<0.001$) in viability between the highest concentration (121.5 μM) and the solvent control. This solvent "shadow" effect may be circumvented by preparing a more concentrated stock solution; for example, 2 mg/ml instead of 1 mg/ml. As for the HepG2 cells, we did not observe a significant difference ($P<0.001$) in viability between the untreated control and the solvent control, while there was a significant difference ($P<0.001$) between the highest concentration (121.5 μM) and the solvent control.

Even though the results for HeLa treatment with our peptide drug did not yield any reliable dose response curve, we were in fact able to observe morphological differences between treated and untreated cells. The cells treated with 121.5 μM displayed numerous large vacuoles indicative of cellular distress. It would be of importance to try more time points for the treatment and observe if increasing treatment duration would result in a dose response to the peptide drug treatment.

In this study, we were able to construct an SVM model in order to sift through the AUC Red Sea Metagenomic library and identify potential anticancer peptides. Out of a list of about 59 potential hits, we were able to isolate one peptide using cationicity, length, and model performance score as the main criteria. We were also able to optimize this peptide for better model performance. 3D modelling and sequence/ structure alignments provided insight into the potential action of the selected peptide. Even though we were not able to clearly propose a distinct mechanism of action for our peptide, we were able to observe a dose- dependent cytotoxicity upon treatment with our peptide. In addition, we observed that our peptide displayed less toxicity towards normal cells than cancer cells. The performance and selectivity of our peptide can be augmented by introducing more mutations into the amino acid sequence. More in-depth experimentation is needed, however, in order to outline a more pronounced hypothesis regarding the mechanism of action of our peptide drug. The activity of caspases 3 and 9 along with Apaf-1 need to be examined. Furthermore, investigation of the mitochondrial potential may prove beneficial in order to investigate whether the peptide drug is directly disrupting the mitochondria. Our peptide drug may also be tested in combination with chemotherapeutic agents in the hope that the additive effect of both agents would lower the required dosage of each and reach a point of very low toxicity to normal cells. Towards evaluating toxicity, further testing would be required in order to shed light on levels of hemolytic activity of our peptide drug. There still remains much research to be carried out in the field of cancer therapeutics. The need for more specific, more potent, and less toxic therapeutics is higher than ever due to the increasing global cancer burden.

REFERENCES

1. Cancer, W.H.O.I.A.f.R.o., *World Cancer Report*, in *World Cancer Report 2014*, C.P.W. Bernard W. Stewart, Editor. 2014, World Health Organisation.
2. Institute, N.C. *What is Cancer*. 2016 February 9th, 2015 [cited 2016 November 23rd]; Available from: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.
3. Weinberg, R.A., *The Nature of Cancer*, in *The Biology of Cancer*. 2007, Garland Science: New York, United States of America. p. 25- 56.
4. Thundimadathil, J., *Cancer treatment using peptides: current therapies and future prospects.* J Amino Acids, 2012. **2012**: p. 967347.
5. Schweizer, F., *Cationic amphiphilic peptides with cancer-selective toxicity.* Eur J Pharmacol, 2009. **625**(1-3): p. 190-4.
6. Craik, D.J., et al., *The future of peptide-based drugs*, in *Chem Biol Drug Des*. 2013. p. 136-47.
7. Riedl, S., D. Zweytick, and K. Lohner, *Membrane-active host defense peptides--challenges and perspectives for the development of novel anticancer drugs.* Chem Phys Lipids, 2011. **164**(8): p. 766-81.
8. Wu, D., et al., *Peptide-based cancer therapy: opportunity and challenge.* Cancer Lett, 2014. **351**(1): p. 13-22.
9. Mulder, K.C., et al., *Current scenario of peptide-based drugs: the key roles of cationic antitumor and antiviral peptides.* Front Microbiol, 2013. **4**: p. 321.
10. Zhao, J., et al., *In vitro Characterization of the Rapid Cytotoxicity of Anticancer Peptide HPRP-A2 through Membrane Destruction and Intracellular Mechanism against Gastric Cancer Cell Lines.* PLoS ONE, 2015. **10**(10): p. 1-11.
11. Hamannn, M.T. and C.S. Otto, *Kahalalides: Bioactive peptides from a marine mollusk Elysia rufescens and its algal diet.* Journal of Organic Chemistry, 1996. **61**(19): p. 6594.
12. Sinko, J., et al., *Biologically active substances from water invertebrates: a review.* Veterinární Medicína, 2012. **57**(4): p. 177-184.
13. Wang, C., et al., *Melittin, a Major Component of Bee Venom, Sensitizes Human Hepatocellular Carcinoma Cells to Tumor Necrosis Factor-related Apoptosis-inducing Ligand (TRAIL)-induced Apoptosis by Activating CaMKII-TAK1-JNK/p38 and Inhibiting IκBα Kinase-NFκB.* Journal of Biological Chemistry, 2009. **284**(6): p. 3804-3813.
14. Tyagi, A., et al., *In silico models for designing and discovering novel anticancer peptides.* Sci Rep, 2013. **3**: p. 2984.
15. Wang, G., X. Li, and Z. Wang, *APD2: the updated antimicrobial peptide database and its application in peptide design.* Nucleic Acids Res, 2009. **37**(Database issue): p. D933-7.
16. Wang, Z. and G. Wang, *APD: the Antimicrobial Peptide Database.* Nucleic Acids Research, 2004. **32**(suppl 1): p. D590-D592.
17. Thomas, S., et al., *CAMP: a useful resource for research on antimicrobial peptides.* Nucleic Acids Research, 2010. **38**(suppl 1): p. D774-D780.

18. Novković, M., et al., *DADP: the database of anuran defense peptides.* Bioinformatics, 2012. **28**(10): p. 1406-1407.
19. Waghu, F.H., et al., *CAMP: Collection of sequences and structures of antimicrobial peptides.* Nucleic Acids Research, 2014. **42**(D1): p. D1154-D1158.
20. Eddy, S.R., *Profile hidden Markov models.* Bioinformatics, 1998. **14**(9): p. 755-763.
21. Finn, R.D., et al., *Pfam: the protein families database.* Nucleic Acids Research, 2014. **42**(D1): p. D222-D230.
22. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nat Protoc, 2010. **5**(4): p. 725-38.
23. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
24. Yang, J., A. Roy, and Y. Zhang, *Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment.* Bioinformatics, 2013. **29**(20): p. 2588-95.
25. Yang, J., A. Roy, and Y. Zhang, *BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1096-103.
26. Knowles, B., C. Howe, and D. Aden, *Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen.* Science, 1980. **209**(4455): p. 497-499.
27. Masters, J.R., *HeLa cells 50 years on: the good, the bad and the ugly.* Nat Rev Cancer, 2002. **2**(4): p. 315-319.
28. Rahbari, R., et al., *A novel L1 retrotransposon marker for HeLa cell line identification.* BioTechniques, 2009. **46**(4): p. 277-284.
29. Ziko, L., et al., *Mechanical stress promotes cisplatin-induced hepatocellular carcinoma cell death.* Biomed Res Int, 2015. **2015**: p. 430569.
30. Ormerod, M.G., et al., *Increased membrane permeability of apoptotic thymocytes: A flow cytometric study.* Cytometry, 1993. **14**(6): p. 595-602.
31. Andree, H.A., et al., *Binding of vascular anticoagulant alpha (VAC alpha) to planar phospholipid bilayers.* Journal of Biological Chemistry, 1990. **265**(9): p. 4923-4928.
32. Lu, J. and Z.-w. Chen, *Isolation, characterization and anti-cancer activity of SK84, a novel glycine-rich antimicrobial peptide from Drosophila virilis.* Peptides, 2010. **31**(1): p. 44-50.
33. Lehane, M.J., D. Wu, and S.M. Lehane, *Midgut-specific immune molecules are produced by the blood-sucking insect Stomoxys calcitrans.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(21): p. 11502-11507.
34. *Family: Homeobox (PF00046)*. [cited 2015 3/8]; Available from: pfam.xfam.org/family/pf00046.24.
35. Vlieghe, P., et al., *Synthetic therapeutic peptides: science and market.* Drug Discov Today, 2010. **15**(1-2): p. 40-56.

36. Trabulo, S., et al., *Cell-Penetrating Peptides—Mechanisms of Cellular Uptake and Generation of Delivery Systems.* Pharmaceuticals, 2010. **3**(4): p. 961-993.
37. Copolovici, D.M., et al., *Cell-Penetrating Peptides: Design, Synthesis, and Applications.* ACS Nano, 2014. **8**(3): p. 1972-1994.

APPENDIX

Youssef Abdou
American University in Cairo
Biotechnology Master's Program
AUC Avenue, P.O. Box 74
New Cairo, 11835 Egypt

Date: 10th Dec. 2016

Dear Youssef,

Mustafa A. Ahmed hereby authorizes Youssef Abdou to include my data related to scanning peptides with potential anticancer / antimicrobial activity on a large scale in metagenomic datasets, illustrated on Red Sea metagenomics library (generated by AUC & KAUST and made public at NCBI SRA) in his master thesis.

Sincerely,

Mustafa A. Ahmed