

Identification and In-silico Prediction of Hermetiaillucens Larval Protein that Inhibits the Progression of Cervical Cancer

Nisha Rajasekhar, Ramesh N, Prashantha C.N.

Abstract: The aim of the current research is to assess the anticancer activity of the proteins identified from the crude protein extract derived from *Hermetiaillucens* larvae by computational methods. Bioinformatics methods were used to identify the protein sequence and virtual screening for the prediction of protein structure, their physico-chemical characteristics and functional aspects, that aid in exploring the anti-cancer interaction that is inclined to inhibit the oncoprotein activity. Due to the pathological process transpired by HPV, an array of genetic alterations, including the overexpression of oncoproteins or the inactivation of tumor suppressor proteins like TP53, BCL2, MDM2, ARF and BAX occur. The altered protein signaling pathways convulse to cervical cancer. In this research, we identified proteins, from the crude extract derived from *Hermetiaillucens* by LC-MS method. Further Computational screening of the selected proteins were employed to assess the functional units which has anticancer activity. We predicted four proteins Metallothionine, Defensin like precursor 1 protein, Heat shock protein 90 and NADH dehydrogenase and eluted the protein sequence from NCBI GenBank database. These sequences were used to predict physicochemical properties and protein structure prediction. Pharmacophore analysis of the peptide sequences as potential targets for cancer treatment was evaluated. Molecular docking of peptide sequences with target protein structure was carried out. To screen the best active potential molecule for cancer treatment the Molecular dynamics of docked protein-peptide structures were administered. The molecular docking of the peptides with onco-proteins has been predicted and virtually screened based on RMSD values.

The resultant protein sequence was evaluated further by amino acid sequencing, extensive scrutiny of the proteins, advanced computational imaging, molecular docking examination and assessed the results for clinical interpretation. To summate, bioinformatics assessment enabled us to identify four key proteins derived from *Hermetiaillucens* larvae that have strong inhibitory function against proteins that cause cervical cancer.

Keywords: Pharmacokinetics Cell proliferation, LC-MS, quantification of protein, Bioinformatics assessment, identification and characterization, Bioactive prediction, Peptide Ranking , Target Protein selection, Molecular docking.

I. INTRODUCTION

Cancer is a genetically modified uncontrolled cell proliferation that has multistep process arising from gene mutations involved in signaling, cell-cycle and/or cell-death pathways resulting in mis-regulation.[1]. Cervical cancer is the fourth most common cancer in women worldwide, with

Revised Manuscript Received on February 05, 2020.

Nisha Rajasekhar: Research Scholar, Department of Biotechnology, REVA University, Bangalore-560064.

Ramesh N, Director, Dean-Training, Placement and planning, School of Applied Sciences, REVA University, Bangalore-560064.

Prashantha C.N, Assistant Professor, Department of Biotechnology, REVA University, Bangalore-560064.

an estimated 527,600 new cases and 265,600 deaths in 2012[2]. Tumorigenesis is a complex pathological process involving a variety of genetic alterations, including the over expression of oncogenes and/or the inactivation of tumor suppressor genes. Either way, the cell cycle is disrupted leading to uncontrolled cell division and growth -a common feature of every cancer type [3]. The development of cervical cancer is a stepwise process from a low-grade cervical intraepithelial neoplasia that ultimately progresses into carcinoma involving multiple genetic and epigenetic events[4]. The identification of dysregulated genes in cancer-associated pathways may shed light on the molecular mechanisms underlying tumorigenesis, thus helping to develop new strategies for tumor therapy [5]. There are 55% of women who is affected every year; in the year 2018 over 570,000 new cases were diagnosed in cervical cancer [6]. It is found that 17% of all cancer deaths among women occurs between thirty and sixty nine years and approximately 1 in 53 Indian women suffer from Cervical cancer [7]. There is a large demand to control the pathogenesis of HPV in women, which increase to create awareness about the diagnostic and treatment methods to control disease. There are several treatment methods such as surgery, radiation therapy, chemotherapy that has immense advantage for the treatment of cervical cancer. The disadvantage of chemotherapeutic drugs is that it creates side effects by destroying malignant as well as non-malignant cells [8].

There is a great demand to identify the drugs with minimal side effects for the treatment of cervical cancer [9,10]. Protein therapy approach would be a new and effective treatment of cancer without causing any side effects unlike chemotherapy. Protein has the capability to effectively destroy cells which are linked to the progression and development of cancer. Peptides have tremendous application in the field of Biotechnology and medicine field. At present more than 60 peptide-based drugs are available in the US Market which has been approved by FDA. [11]. Peptide drug discovery has secured scientific propulsion as it confronts many unidentified medical concerns.[12]

The current research is focused on the development of a peptide-based drug that aid to control cervical cancer. Here, the execution of Bioinformatics in cancer research revamped the next generation protein sequencing and has tailored our interpretation of cancer biology approaches. In this research we are focusing on the alternative and application specific approaches that are now becoming apparent to improve a wide variety of prevailing analytic techniques.

II. MATERIALS AND METHODS

A. Quantification of Protein-peptides using LC-MS

Trypsin digested sample was passed through RP C18 column 4.6 x 150mm with the linear gradient starting from 5% of ACN to 95% ACN in 50 Minutes. Mobile phase used was water and Acetonitrile with 0.1% FA. The LC gradient applied allows the separation of major proteins present in *Hermetiaillucens* and showcased the results. The LC-MS spectral results and comparison of results with library search has successfully enabled the identification of the total number of proteins present in the crude sample. Several protein hits were observed from NIST and MASCOT database. The full length of protein sequences was retrieved from the NCBI database to predict further functional characters that can be treated with cancer proteins

Bioinformatics Analysis: Peptide design and Molecular mass prediction

The current technological development in the field of Molecular Biology have extensively provided in contributing the scientific research associates with enormous quantity of Biological data base information [13]. The analogue between the genes and protein play a prominent role in the probe of cancer molecular mechanism. It is imperative to integrate the Cancer research with biology, bioinformatics and computational science to revamp cancer therapy [14]. The enormous output of diverse protein/genome sequencing strategy have administered advancement in the protein therapeutics in a remarkable scale. The protein peptide synergy is accepted as a channel to contribute functional specificity and control the activity of molecules and signaling pathways [15]. In-silico drug modeling dwells with the assemblage of mechanism which aloe to make analytical decisions with respect to description/identification of biomolecular target of therapeutic significance, choosing of new peptides and their transformation to acquire exceptional inclination and pharmacokinetic and pharmacodynamics characteristics[16]

Based on the prediction of peptide sequences from LC-MS analysis, further to predict the functional peptide sequences using *NIST databases*. The resultant peptide sequences were tested with in-silico analysis to screen anticancer peptides. Here, we have used various approaches to screen the lowest toxicity and minimum side effects using computational approaches. There are several bioinformatics steps used to screen peptides which are mentioned below.

B. Selecting the suitable enzymes and protein sequences

Using LC-MS analysis to identify the proteins present in the crude extract, the *NCBIGenBank* database is used to download the full length of the protein sequences and used to predict functional peptides. *Peptide cutter* is a potential proteomic tool to predict the potential cleavage sites cleaved by protease enzymes and chemical digestion with possible cleavage sites. With the employment of *Enzyme-Predictor* mechanism which expeditiously checks the enzymes that effectively cleaves the sample including peptides from the hydrolyzed proteins. This mechanism develops a colorful visualization of the source proteins, the hydrolysate, the fragments, and the enzymes that are predicted and cleaves these fragments.

C. Identification and characterization of peptide sequences

Employed **BIOPEP**, **PeptideDB**, **CAMP**, **APD2**, and **PepBank** software's to characterize the peptide sequences based on biological process, molecular function and cellular activity of each peptide sequence or amino acid which expressed the most convincing effect of anticancer and antioxidant properties. The BioPEP and Peptide Cutter helped to specify the protein substrates of enzymatic hydrolysis to specify enzymatic activity. Subsequently we utilized virtual digestion in combination with insect protein-peptide that specifies the Physico-chemical characteristics such as molecular weight, theoretical PI, stability of the peptides, which enable to predict the potential bioactive peptides. There are several multiple enzyme factors that influence cancer inhibition which needs to be identified by using both in-vitro and in-vivo studies.

D. Bioactive prediction of peptides

Peptide Ranker tool was employed to predict the bioactive property of peptide sequences and is compared with reference target sequences to understand the comparison between target and template sequences. The peptide structures were predicted and modeled using Pep-Fold server by taking into consideration of the least energy model of peptide stability. **Argus Lab** anticipates the peptide poses and facilitates refinement by minimization of energy and/or simulations of molecular dynamics.

E. Target protein selection

Based on the cytotoxicity study, the crude protein is treated with MCF7 and HeLa cells using MTT assay method. The crude protein showed more growth inhibition in the HeLa cell lines with respect to MCF7. Further we examined the anti-apoptotic and cell cycle inhibitory study and the fact showed that the inhibitory effect in apoptosis has eventuated in the G1/S phase and this facilitated us to determine the protein sequences and structures from *NCBI-GenBank*, *UniProtKB* and *PDB* databases (<http://www.rcsb.org/pdb>).

The molecular analysis of apoptotic signaling pathways unveiled various target receptors that help to regulate and inhibit several proteins. Predominantly TP53, MDM2, BCL2, ARF, and BAX signaling proteins are involved in apoptosis [15]. The selected protein sequences were used to predict sequence similarity to find templates by *PSI-BLAST*. The selected template sequences were used to build three-dimensional protein structures using *Swiss model*. Adopting homology modeling by *SwissPDBViewer* aided us to predict the complexity of the protein structures. The protein structures were validated to check the overall protein quality and stereo-chemical activity of atoms and amino acids. The validation was predicted by *Structural analysis and verification server (SAVES)*. The conformational complexity of protein structures was utilized to predict active site amino acids that help for ligand binding using *CastP* calculation server and the best complex protein structures were used for molecular docking.

F. Molecular Docking

Molecular docking is a significant mechanism which contributes identification, optimization and anticipation of the inherent position, orientation and arrangement of a small molecule within the binding site of a targeted molecule. High throughput screening is resorted at the initial stage of the drug interaction process for the verification of numerous compounds for potential interaction against the chosen target [17-21]

Target protein and peptide sequences were docked using *Cabs-Dock* and Galaxy *Pep Dock* servers. This provides fully automated template-based docking of selected peptides by protein-peptide interaction studies. The protein sequences were built by the complex structures using energy-based optimization and refinement that allows structural flexibility. Template-based docking of highly homologous complexes is also provided by protocols dedicated to the prediction and design of peptide binding specificity. The *Flex Pep Bind tool* is used to model the different protein peptides and binding interactions with receptor binding sites with the constraint of hydrogen bonds were predicted that shows the conserved hydrogen bonding. The virtual screenings of docked peptide structures were used for molecular dynamics methods to calculate the electrostatic and Vander Waals interactions to understand the potential surface energy of the peptide-protein interactions. The particular resultant structures will help to screen further in-vitro and in-vivo analysis for further clinical applications for the treatment of cancer.

III. RESULTS

G. LC-MS-Data of probability Based on Mowse Score

We conducted the LC-MS analysis of the crude protein to identify the proteins present in the crude sample. Utilized MASCOT database to identify the protein-peptide hits present in the LC-MS data to predict the number of peptide sequences present compared with the reference protein databases such as UniProt and NCBI biological databases. Further employed NIST software to identify the protein peptides present, based on the molecular weight prediction. Subsequently few protein hits were observed from NIST and MASCOT database. The full length of protein sequences was retrieved from the NCBI database to predict further functional characters that can be treated with cancer proteins. Proteins identified from LC-MS analysis are listed in (**Table: 1**). Implemented bioinformatics methods to understand the protein-peptide regions based on functional activities. InterproScan was used to predict the functional domains and profiles that can characterize functional enrichments of protein sequences. Based on InterproScan analysis, we identified that some of the functional profiles of protein-peptide that have anticancer activities (**Table: 2**).

The peptide character prediction was analyzed using PEPBUB and identified that anticancer activity was seen in Metallothionein, Defensin-like peptide 1, Heat shock protein 90 and NADH dehydrogenase and these proteins were selected for further studies (**Figure: 1**). Bioinformatics tool *Peptide Cutter* was utilized based on chymotrypsin digestion default option that helps to cleave the proteins with different peptides and further to screen biological

inhibitory activity using PEPBUB tool (**Table: 3**). Peptides were ranked based on biological protein inhibition activity which exhibited the best active peptide sequences for antioxidant and anticancer activity. The Physico-chemical properties of the selected active peptides with significant inhibition shows instability index of 37-40 (stable) and >40 (unstable) properties (**Table: 4**).

H. Peptide ranking

For each protein in the query, the ranked list of tryptic peptides can be browsed by clicking and expanding the protein row. Peptides uniquely present in the database (at the protein level) are highlighted. We utilized the search box to search for a specific protein name or peptide sequence. The outcome revealed that 2 peptide sequences, Metallothionein and NADH bioactive peptide sequences are unstable in nature, and Defensin-like peptide 1 and heat shock protein sequences exhibited stable nature. The Peptide ranking of selected peptides are conducted based on bioactive prediction and ranking (**Table: 5**). Bioinformatic tool *PEP-FOLD* was employed to predict the three-dimensional protein structure using the server, assisted to generate the 3D complexity of protein structure (**Figure: 2**).

Prediction of the homology models by employing Swiss Model and energy minimization by utilizing SwissPDB Viewer software (**Table: 6**). The quality analysis of proteins was predicted using SAVES to check the stereo-chemical activity of amino acids arranged in a complex protein structure by using PROCHECK, ERRAT, Verify3D parameters that facilitated to understand the complexity of the protein structure. Active site amino acids were predicted using PDBSum and CastP calculation servers to recognize the peptide or ligand binding sites of TP53, BCL2, MDM2, ARF and BAX protein based on surface area and surface volume (**Figure:3**).

The protein-peptide interactions were investigated using Cabs Dock online server to understand the target protein and peptide sequences. The target proteins were docked with Metallothionein peptide sequence and the interaction of the protein peptides were screened based on cluster RMSD values. The cluster RMSD values which exhibited highest value has strong interaction with highest ranking. Protein-peptide docking of Metallothionein sequence with target protein structures (**Figure: 4**) Molecular docking of Defensin like Protein-1 docked with the target proteins TP53, BCL2, MDM2, ARF, BAX. The Defensin-like peptide 1 is docked with target proteins and the output indicated BCL2 (40.0659±32.534 kcal/mol), BAX (31.8165±23.260 kcal/mol) and MDM2 (21.3619±21.116 kcal/mol) proteins exhibit strong interactions compared with reference drug compound (Camptothecin). (**Figure:5**) Heat Shock Protein 90 docked with the target proteins TP53, BCL2, MDM2, ARF, and BAX. Heat shock protein-peptide sequence is docked with target proteins and the resultant protein structures are screened based on scoring of RMSD values BAX (32.256±23.260 kcal/mol), BCL2 (26.8971±32.534 kcal/mol), TP53 (23.7712±98.531 kcal/mol) MDM2 (19.8872±21.116 kcal/mol). This protein expressed strong interaction compared with reference drug compound.

Identification and In-silico Prediction of Hermetiaillucens Larval Protein that Inhibits the Progression of Cervical Cancer

(Figure:6). NADH dehydrogenase docked with the target proteins TP53, BCL2, MDM2, ARF, and BAX. The NADH dehydrogenase protein when docked with target proteins and the interaction scores were screened based on RMSD values and the resultant structures manifested that MDM2 (42.5102 ± 21.116 kcal/mol), ARF (26.8584 ± 104.57 kcal/mol), BAX (24.8126 ± 23.260 kcal/mol) and BCL2 (22.7283 ± 32.534 kcal/mol) proteins show strong interactions compared with reference drug compound (Camptothecin) (Figure: 7). Reference drug Camptothecin docked with the target proteins TP53, BCL2, MDM2, ARF, BAX. Molecular docking assessment of reference Camptothecin drug with target proteins are showcased for better understanding of the activity the predicted peptides. (Figure 8).

IV. DISCUSSION

Research on Protein/peptides and therapeutic analysis paved a way to understand the mechanism of molecular interactions with therapeutic applications. In the current research we utilized the proteins extracted from *Hermetiaillucens* larvae, which has great precedence in multiple biological functions such as anti-oxidant and anticancer properties. We conducted the LC-MS analysis of the crude protein to identify the proteins present in the crude sample. Utilized MASCOT database to identify the protein-peptide hits present in the LC-MS data to predict the number of peptide sequences present compared with the reference protein databases such as UniProt and NCBI biological databases. Further employed NIST software to identify the protein peptides present, based on the molecular weight prediction. Subsequently 20 protein hits were observed from NIST and MASCOT database. The full length of protein sequences was retrieved from the NCBI database to predict further functional characters that can be treated with cancer proteins.

We then proceeded with implementing bioinformatics methods to understand the protein-peptide regions based on functional activities. InterProScan was used to predict the functional domains and profiles that can characterize functional enrichments of protein sequences. Based on InterProScan analysis, we identified that some of the functional profiles of protein-peptide have anticancer activities. The peptide character prediction was analyzed using PEPBUB and identified that anticancer activity was seen in Metallothionein, Defensin-like peptide 1, Heat shock protein 90 and NADH dehydrogenase and these proteins were selected for further research. Subsequently we employed *Peptide Cutter* based on chymotrypsin digestion default option that helps to cleave the proteins with different peptides and further to screen biological inhibitory activity using PEPBUB tool. Peptides were ranked based on biological protein inhibition activity which exhibited the best active peptide sequences for anticancer activity. The Physico-chemical properties of the selected active peptides with significant inhibition shows in stability index of 37-40 (stable) and >40 (unstable) properties. The outcome revealed that 2 peptide sequences, Metallothionein and NADH bioactive peptide sequences are unstable in nature and Defensin-like peptide 1 and heat shock protein

sequences exhibited stable nature. Eventually PEP-FOLD was employed to predict the three-dimensional protein structure using the server, assisted to generate the 3D complexity of protein structure.

As per the outcome, inhibitory study of crude protein with HeLa cells revealed strong inhibition on cancer cells which steered the uncovering of the target proteins that counter cervical cancer. Eventually KEGG database was linked to find the pathway of apoptotic proteins involved in cervical cancer based on viral infection. Here, we selected TP53, BCL2, MDM2, BAX and ARF target protein for the analysis. The protein sequences were retrieved from UniProtKB database and structural similarity was predicted to understand the similar templates using PSI-BLAST. We then predicted the homology models by employing Swiss Model and energy minimization by utilizing Swiss PDB Viewer software.

The quality analysis of proteins was predicted using SAVES to check the stereo-chemical activity of amino acids arranged in a complex protein structure by using PROCHECK, ERRAT, Verify3D parameters that facilitated to understand the complexity of the protein structure. Active site amino acids were predicted using PDBSum and CastP calculation servers to recognize the peptide or ligand binding sites based on surface area and surface volume. The protein-peptide interactions were investigated using Cabs Dock online server to understand the target protein and peptide sequences. The target proteins were docked with Metallothionein peptide sequence and the interaction of the protein peptides were screened based on cluster RMSD values. The cluster RMSD values which exhibited highest value has strong interaction with highest ranking.

The results show that TP53 and BAX has strong interaction with Metallothionein peptide sequence compared with reference drug compound (Camptothecin) The Defensin-like peptide 1 is docked with target proteins and the output indicated BCL2 BAX and MDM2 proteins exhibit strong interactions compared with reference drug compound (Camptothecin). When Heat shock protein-peptide sequence is docked with reference proteins values BAX, BCL2, TP53, MDM2, expressed strong interaction compared with reference drug compound. The NADH dehydrogenase protein when docked with target proteins and the interaction scores were screened based on RMSD values and the resultant structures manifested that MDM2, ARF, BAX and BCL2 proteins show strong interactions compared with reference drug compound (Camptothecin). Molecular docking assessment of reference Camptothecin drug with target proteins are showcased for better understanding of the activity the predicted peptides.

V. CONCLUSION

Peptides are considered to possess high clinical therapeutic value. Peptides are significant in the therapeutic field as it can be customized to achieve desired biocompatibility that brands it to a potent therapeutic grade. Currently peptide therapeutics is a major research category in the field of identification of peptide sequences for the treatment of cancer and other diseases. In cancer therapeutics,

peptides demonstrate a significant role because of its high target specificity and selectivity, high biological and chemical diversity, high solubility, diminutive toxicity and inconsequential side effects.

The execution of Bioinformatics in cancer research revamped the next generation protein sequencing and has tailored our interpretation of cancer biology approaches. In this research we are focusing on the alternative and application specific approaches that are now becoming apparent to improve a wide variety of prevailing analytic techniques. In this research we resorted to *Hermetiaillucens* insect larvae protein and analyzed that the peptides offer enormous potential for anticancer activities. We adopted Bioinformatics assessment to understand the protein peptides that are present in the LC-MS data and also aided to screen the peptides based on the interaction with apoptotic proteins. In this study we predicted four pre-eminent peptide sequences that interacted and bind with target proteins and the outcome portrayed strong interaction with TP53, BCL2, MDM2, ARF and BAX proteins with highest RMSD and Cluster Maps. The outcome exhibited the prime functional peptide unit sequence from Metallothionein (CASQTQSSGSCGCGSYCK), Defensin-like peptide 1- precursor (QQGWAMGDCR), Heat shock protein 90 (ADLVNNLGTIAK) and NADH dehydrogenase (GIPQPFCDVAVKLFTKE) shows strong inhibition with apoptotic proteins. The conclusion provides great scope for experimental validation in utilizing *In-vitro* and *in-vivo* studies, by synthesizing individual peptide sequences or isolate individual peptide sequences from *Hermetiaillucens* crude protein extract. These experimental studies would certainly help validate the therapeutic function of peptides for the utilization of cervical cancer treatment and aid its clinical relevance.

Table 1: Sequence identification and retrieval from NCBI GenBank Database (functional unit)

Protein	Peptide sequence
Metallothionein	MPCPCGTGCKCASQTQSSG SCGCGSYCKGSSCDKGS
Defensin-like peptide 1	CIGNGGHCQPDGSMGNCC SGNCYKQQGWAMGDCR
Heat shock protein 90	VETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNS SDALDK IRYESLTDPSKLDGKELYIKIIPNKAAGTLTIIDT GIGMTKA DLVNNLGTIAKSGTKAFMEALQAGADISMIGQF GVGFYSAY LVADKVTVTSKHNDDEQYTWESAAGGSFTVSN DHSEPLGRG TKIVLHIKEDQMEYLEEAKINSIVTKYSQFIRYPI KLLVKKER EKEVSDDEAE
NADH dehydrogenase	GIPQPFCDVAVKLFTKE

Table 2: Protein functional region identification using InterProScan

Protein	Active position	Function
Metallothionein	1-37	Metal ion binding
Defensin-like peptide 1	58-98 28-62 30-62	Defensin-like peptide 1_invertebrate/Antifungal/Antibacterial

Heatshock protein inhibitor 90	9-213 279-534 534-683 31-185 29-38 9-29	HATPase_C_sf Ribosomal_S5_D2-typ_fold HSP90_C HATPase_C Heat_shock_protein_90_CS Hsp90_N
NADH dehydrogenase subunit 1	51-66, 204-217 14-303,	NADH_UbQ_OxRdtase_su1/FPO

Table 3: Peptide character prediction using PEPBUB bioinformatics tools

Protein	Peptide sequence	Inhibitory function	Functional regions
Metallothionein	MPCPCGTGCKCASQTQSSG SCGCGSYCKGSSCDKGS	ACE inhibitor	6, 7, 18, 19, 20, 24, 25, 30, 31, 35-36
		dipeptidyl peptidase IV inhibitor	1, 2, 12, 13, 14, 15, 16, 17,25, 26, 35, 36,
Defensin-like peptide 1-like peptide 1	CIGNGGHCQPDGSMGNCC SGNCYKQQGWAMGDCR	Peptide activating ubiquitin-mediated proteolysis	28-29
HSP 90 inhibitor	VETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNS SDALDK IRYESLTDPSKLDGKELYIKIIPNKAAGTLTIIDT GIGMTKA DLVNNLGTIAKSGTKAFMEALQAGADISMIGQF GVGFYSAY LVADKVTVTSKHNDDEQYTWESAAGGSFTVSN DHSEPLGRG TKIVLHIKEDQMEYLEEAKINSIVTKYSQFIRYPI KLLVKKER EKEVSDDEAE	Stimulating vasoactive substance release	183-184 218-220
		Immunostimulating peptide Neuropeptide Anxiolytic peptide Antioxidative	116, 117 126, 127, 181, 182 172-173 31-32,43, 44, 59-60, 198, 199
NADH dehydrogenase	GIPQPFCDVAVKLFTKE	Antioxidative peptide	6, 7, 10, 11, 12,

Table 4: Physicochemical property prediction by utilizing ProtParam

Protein	Number of Amino acids	Molecular Weight	Theoretical pI	Stability	Hydropathicity
Metallothionein	37	3560.02	8.11	Unstable (61.84)	-0.300
Defensin-like peptide 1-like peptide 1	34	3538.93	6.70	Stable (12.80)	-0.679

Identification and In-silico Prediction of Hermetiaillucens Larval Protein that Inhibits the Progression of Cervical Cancer

HSP 90 inhibitor	220	24410.16	4.94	Stable (20.57)	-0.235
NADH dehydrogenase	16	1793.11	6.06	Unstable (43.31)	-0.013

Table 5: Peptide ranking of the selected peptides based on bioactive prediction and ranking.

Protein	Peptide sequence	Position	Length	Score	Rank
Metallothionein	CASQTQSSGSCGCGSYCK	11	18	-1.473	1
	MPCPCGTGCK	1	10	-6.348	2
	CGSSCDK	29	7	-9.922	3
Defensin-like peptide 1-like peptide 1	QQGWAMGDCR	25	10	-3.407	1
	CIGNGGHCQPDGSMGNCCSGNICYK	1	24	-4.239	2
Heat Shock protein 90	HNDDEQYTWESAA GGSFVSNHDHSEPLGR	138	29	-1.308	1
	ADLVNNLGTIAK	85	12	-1.397	2
	AAGTLTIIDTGIGMTK	69	16	-2.116	3
	ELISNSSDALDK	31	12	-2.177	4
	EDQMEYLEEAK	176	11	-2.811	5
NADH dehydrogenase	GIPQPFCDAVKLFTKE	1	11	-2.548	1

Table 6: Template search (PSI Blast) and Homology modeling (Swiss Modelling)

Protein	Template	Identity	E-value	Model quality	Active sites (AA)	Verify 3D	Quality
TP53	4LO9	99.54	0.007	1.57%	Cys238, His179, Cys176, Cys242	100%	87.97
BCL2	1GJH	90.96	0.008	-0.687%	Asp3, Lys5, Lys6, Leu8, Asp9, Lys12, Asp13, Arg92, Glu95, Glu99,	84.39%	71.42
MDM2	2LZG	100.00	0.0037	-0.37%	His96, Thr16, Val14, Leu54, Gly58, Ile61, Val93, Tyr67, Lys94	100%	97.36
ARF	1HUR	100.00	0.000	-0.000%	Ala27, Ala28, Gly29, Lys30, Thr31, Thr32, Asn52, Gly50, Asn126, Ala160, Lys127, Asp129, Leu130,	86.59%	91.66

						Thr161,		
BAX	1F16	100	0.0	-4.36	89.2%	Leu63, Ile66, Gly67, Leu70, Leu76, Gln77, Ile80, Ala81, Val83, Asp84, Thr85, Pro88, Val91, Phe92, Val111, Ala112, Tyr115, Phe116, Leu120, Lys123, Ala124, Thr127, Leu132, Thr139, Val177, Leu181, Trp188	86.98%	87.71

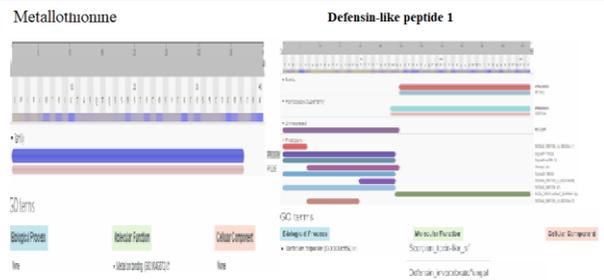


Figure 1: Protein profiles from selected crude sample
Figure Legend: The peptide character prediction analyzed using PEPBUB (Figure: 1).

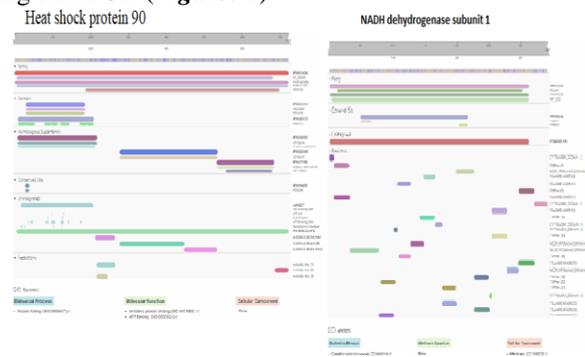


Figure 2: Peptide structure prediction using PEPFOLD Server (<http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD>)

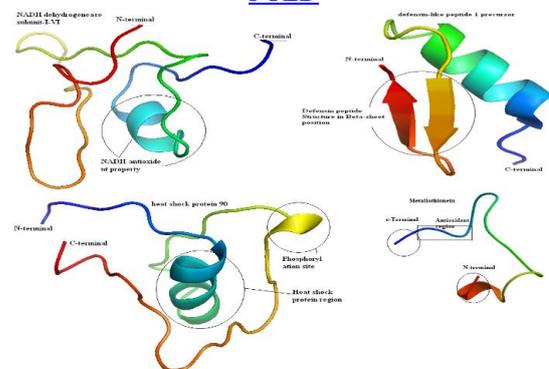


Figure 3: Target protein structure of TP53, BCL2, MDM2, ARF and BAX predicted using Swiss Modelling

Figure Legend: Peptide or ligand binding sites of TP53, BCL2, MDM2, ARF and BAX protein based on surface area and surface volume. (Figure: 3).

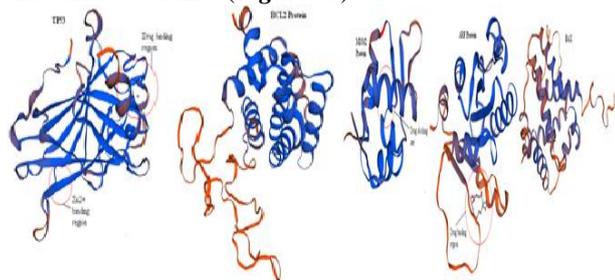


Figure 4: Protein-peptide docking of Metallothionein sequence with target protein structures.

Figure Legend: Molecular docking analysis of Metallothionein with the target proteins TP53, BCL2, MDM2, ARF, and BAX. (Figure: 4)

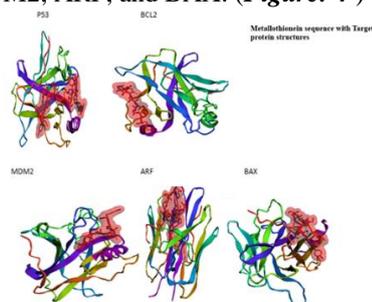


Figure 5: Protein-peptide docking of Defensin-like protein 1 sequence with target protein structures.

Figure Legend: Molecular docking analysis of Defensin like Protein-1 with the target proteins TP53, BCL2, MDM2, ARF, and BAX. (Figure:5)

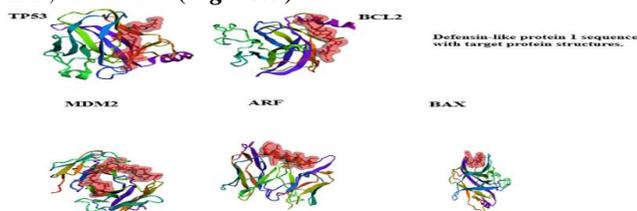


Figure 6: Protein-peptide docking of Heat shock protein 90 sequences with target protein structures.

Figure Legend: Molecular docking analysis of Heat Shock Protein 90 with the target proteins TP53, BCL2, MDM2, ARF, and BAX. (Figure:6).

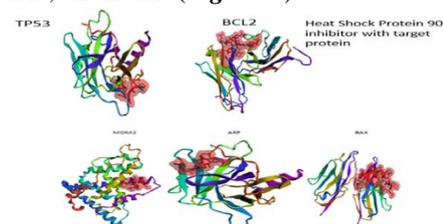


Figure 7: Protein-peptide docking of NADH dehydrogenase protein sequence with target protein structures.

Figure Legend: Molecular docking analysis of NADH dehydrogenase with the target proteins TP53, BCL2, MDM2, ARF, and BAX.

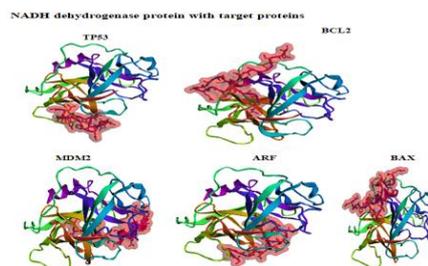


Figure:8. Molecular docking of reference drug Camptothecin with target proteins

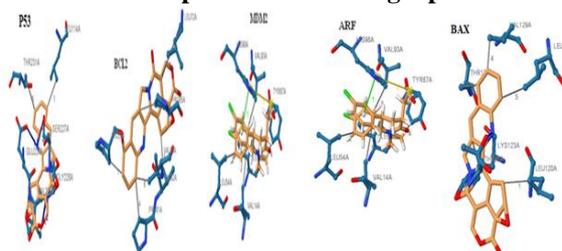


Figure Legend: Reference drug Camptothecin docked with the target proteins TP53, BCL2, MDM2, ARF, and BAX. (Figure 8).

Tables

1. Table 1: Sequence identification and retrieval from NCBI GenBank Database(functional unit)
2. Table 2: Protein functional region identification using InterproScan
3. Table 3: Peptide character prediction using PEPBUB bioinformatics tools
4. Table 4: Physicochemical property prediction by utilizing ProtParam
5. Table 5: Peptide ranking of the selected peptides based on bioactive prediction and ranking
6. Table 6: Template search (PSI Blast) and Homology modeling (Swiss Modelling)

Figures

1. Figure 1: Protein profiles from selected crude sample
2. Figure 2: Peptide structure prediction using PEPFOLD Server
3. Figure 3: Target protein structure predicted using Swiss Model
4. Figure 4: Protein-peptide docking of Metallothionein sequence with target protein structures
5. Figure 5: Protein-peptide docking of Defensin-like protein 1 sequence with target protein structures
6. Figure 6: Protein-peptide docking of Heat shock protein 90 sequences with target protein structures
7. Figure 7: Protein-peptide docking of NADH dehydrogenase protein sequence with target protein structures
8. Figure 8: Molecular docking of reference drug Camptothecin with target proteins

Abbreviations

LC-MS-Liquid Chromatography–Mass Spectrometry, NIST-National Institute of Standards and Technology, NCBI - National Center for Biotechnology Information, BLAST-Basic Local Alignment Search Tool,

Identification and In-silico Prediction of Hermetia illucens Larval Protein that Inhibits the Progression of Cervical Cancer

UniProt – A freely accessible database of protein sequence and functional information, *PDB-Protein Data Bank*, TP53-Tumor Protein 53, MDM2- Mouse Double Minute 2 homolog, BCL2-B-Cell Lymphoma 2, ARF- ADP-Ribosylation factor, BAX- BCL2 Associated X, Apoptosis Regulator, PSI Blast- Position-Specific Iterative, SAVES- Structural Analysis and Verification Server, KEGG- Kyoto Encyclopedia of Genes and Genomes, PROCHECK- a Program to check the stereo-chemical quality of protein structures, ERRAT- web application which intends to assist users in model-building or in structure checking, PDBSUM- web server providing structural information on the entries in the Protein Data Bank, CASTp- Computed Atlas of Surface Topography of proteins.

REFERENCES

1. Tong Ihn Lee and Richard A Young Transcriptional Regulation and its Misregulation in Disease. *Cell*. 2013 Mar 14; 152(6): 1237–1251. doi: 10.1016/j.cell.2013.02.014.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A: Global cancer statistics, 2012. *CA Cancer J Clin* 65: 87-108, 2015.
3. Vogelstein B and Kinzler KW: Cancer genes and the pathways they control. *Nat Med* 10: 789-799, 2004.
4. Wu NL, Huang DY, Tsou HN, Lin YC and Lin WW: Syk mediates IL-17-induced CCL20 expression by targeting Act1-dependent k63-linked ubiquitination of TRAF6. *J Invest Dermatol* 135: 490-498, 2015
5. Kejia Wu, Yuexiong Yi, Fulin Liu, Wanrong Wu, Yurou Chen And Wei Zhang ,Identification of key pathways and genes in the progression of cervical cancer using bioinformatics analysis DOI: 10.3892/ol.2018.8768.
6. Ferlay J, Soerjomataram I, Ervik M, Forman D, Bray F, Dixit R, et al. GLOBOCAN 2012, Cancer Incidence and Mortality Worldwide in 2012. Lyon, France: International Agency for Research on Cancer; 2012. [Last accessed on 2015 Dec 03]. Available from: <http://www.globocan.iarc.fr> . [Google Scholar]
7. Institute for Health Metrics and Evaluation. The Challenge Ahead: Progress in Breast and Cervical Cancer. Institute of Health Metrics and Evaluation. 2011. [Last accessed on 2016 Jan 21]. Available from: <http://www.healthmetricsandevaluation.org/publications/policyreport/challenge-ahead-progress-and-setbacks-breast-and-cervical-cancer>
8. Kulmira Nurgali, R. Thomas Jagoe, and Raquel Abalo, Adverse Effects of Cancer Chemotherapy: Anything New to Improve Tolerance and Reduce Sequelae? *Front Pharmacol*. 2018 Mar 22;9:245. doi: 10.3389/fphar.2018.00245. eCollection 2018.
9. World health organization 2018. Cervical cancer <https://www.who.int/cancer/cervical-cancer>
10. Harvey AL. Natural products in drug discovery. *Drug Discover Today*. 2008; 13:894–901.
11. Keld Fosgerau, Torsten Hoffmann, Peptide therapeutics: current status and future direction <https://doi.org/10.1016/j.drudis.2014.10.003>, *Drug Discover Today*, Volume 20, Issue 1, January 2015, Pages 122-128
12. Henninot A, Collins JC, Nuss JM. The Current State of Peptide Drug Discovery: Back to the Future? *J Med Chem*. 2018 Feb 22;61(4):1382-1414. doi: 10.1021/acs.jmedchem.7b00318. Epub 2017 Aug 11.
13. D. Jawdat, The Era of Bioinformatics ,Date Added to IEEE *Xplore*: 16 October 2006, ISBN: 0-7803-9521-2, INSPEC Accession Number: 9061751, DOI: 10.1109/ICTTA.2006.1684672
14. Duojiao Wu, Catherine M Rice, and Xiangdong Wang Cancer bioinformatics: A new approach to systems clinical medicine, *BMC Bioinformatics*. 2012; 13: 71., Published online 2012 May 1. doi: 10.1186/1471-2105-13-71
15. Ami Aronheim ,Michael Karin ,Analysis and identification of protein-protein interactions using protein recruitment system, *Methods in Enzymology*, Volume 328, 2000, Pages 47-59 ,doi.org/10.1016/S0076-6879(00)28389-4
16. Vincent Zoete ,Aurélien Grosdidier , Olivier Michielin , Docking, virtual high throughput screening and in silico fragment-based drug design, *J. Cell. Mol. Med*. Vol 13, No 2, 2009 pp. 238-248
17. Kontoyianni M, Madhav P, Suchanek E, Seibel W. Theoretical and practical considerations in virtual screening: a beaten field? *Curr Med Chem*. 2008; 15: 107–16.
18. Sousa SF, Fernandes PA, Ramos MJO. Protein-ligand docking: current status and future challenges. *Proteins*. 2006; 65: 15–26.
19. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. 2004; 3: 935–49.
20. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*. 2003; 32: 335–73.
21. Taylor RD, Jewsbury PJ, Essex JW. A review of protein-small molecule docking methods. *J Comput Aided Mol Des*. 2002; 16: 151–66.

AUTHOR PROFILES



Nisha Rajasekhar completed Master of Science and Bachelor of Education from Mahatma Gandhi University, Kerala and has secured degree in Public Relations from National Institute of Human Resource Development, Chennai. Currently she is pursuing PhD in Biotechnology at REVA University, Bangalore. She is a goal-oriented senior administrator with seventeen years of professional experience in various educational organizations.



Dr. N Ramesh is currently adorning the position of Dean-Training and Placement Department, Reva University, Bangalore. He adores a unique place in the continuum of academic administrators, as the facilitating link among Department Chairpersons, School Directors, faculty members, staff, students, and University leadership. He contributes the light of furthering the best interests of the students of the University



Prashantha CN is working as an Assistant Professor in School of Applied Science, Department of Biotechnology, REVA University, Bangalore.