Title : Anticancer Potential of *Eudrilus Eugeniae* Coelomic Fluid Protein on SCC-9 Cells-*In Vitro* Gene Expression entwined with Cutting-Edge *In-Silico* Strategy

Dominic Augustine (dominic2germain@gmail.com)
MS Ramaiah University of Applied Sciences
Roopa S Rao
MS Ramaiah University of Applied Sciences
Hema Sree GNS
MS Ramaiah University of Applied Sciences
Jayaraman Anbu
MS Ramaiah University of Applied Sciences
Saraswathy GR
MS Ramaiah University of Applied Sciences
Chidambara Murthy KN
M.S. Ramaiah Medical College

Research Article

Keywords: Caspase 3, Caspase 8, Molecular Dynamics Simulation, Oligochaeta, Molecular Structure, Molecular Docking Simulation , Matrix-Assisted Laser Desorption-Ionization, Protein Sequence Analysis

Posted Date: April 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-439242/v1

Abstract

Objectives: This study is a pursuit to unearth natural anti-cancer biomolecules from *Eudrilus eugeniae* (EE) Earthworm Coelomic Fluid (ECF), in order to overcome the limitations encountered in pharmacotherapy of oral cancer such as treatment failure and adverse effects. This investigation aims to determine the molecular structure of a18 kDa protein from ECF of EE (18-ECFP) through proteomic and energy-based methods followed by evaluation of its anti-cancer potential through *in silico* and *in vitro* techniques on oral cancer cell line SCC-9.

Materials and Methods: Following advanced proteomic analysis, construction of homology model of 18-ECFP was attempted via energy-based methods using Prime module-Schrodinger. Molecular Dynamic Simulation (MDS) of the modelled 18-ECFP was tailed by Protein-Protein Docking (PPD) with human Caspase-3 and Caspase-8 receptors. The 18-ECFP was further evaluated by RT-PCR and Q-PCR techniques on SCC-9 cells to establish its anti-cancer potential.

Results: The homology model of 18-ECFP exhibited stable molecular dynamics profile. PPD confirmed the binding affinity of 18-ECFP with pro-apoptotic genes Caspase-3 and Caspase-8. Gene expression studies revealed 18-ECFP induced upregulation of pro-apoptotic genes i.e., Caspase-3 and Caspase-8, thus, validating the *in-silico* findings.

Conclusion: This is a first of a kind report to construct a homology model of a novel anti-cancer protein from earthworm source substantiated with MDS and to dock the modelled protein with human caspase receptors supported by multiple gene expression techniques.

Clinical Relevance: The study outcomes may contribute to the development of naturally available drugs to combat oral cancer.

1. Introduction

Oral cancer encompasses a large percentage of head and neck cancers accounting for an estimated 263,000 new cases and about 127,000 deaths worldwide annually. [1] It is the sixth most common cancer in males and the twelfth in females. Oral squamous cell carcinoma (OSCC) is the most frequent histological type of oral cancer representing 90% of total cases. It is a devastating disease with poor prognosis and a 5-year survival rate of less than 50% [2, 3]. Radical surgery and radiation therapy are options for OSCC patients. However, both surgery and radiotherapy often result in unfavorable treatment outcomes in moderate to advanced OSCC cases with adverse impact on quality of life of oral cancer patients [4].

Even with a multimodal approach of surgery, chemotherapy or radiotherapy there is continued proliferation of residual tumor cells resulting in high treatment failure rates. [5] Conventional chemotherapy is a double-edged weapon as it is cytotoxic to both cancerous and normal cells with significant toxicity to haemopoietic cells, lining epithelia of gastrointestinal tract and hair follicles, alongside a deleterious effect on vital organs [5]

To surmount the aforementioned bottleneck constraints in the treatment of oral cancer, current research is oriented towards unraveling alternate therapies with promising therapeutic care alongside an acceptable safety profile. In this regard, the pursuit to unearth natural anti-cancer biomolecules as an adjunct to synthetic chemotherapeutic drugs would be beneficial from a therapeutic standpoint.

The Earthworm Coelomic Fluid (ECF) is one such naturally existing therapeutic option that could well serve this purpose. The ECF is an intestinal secretion that forms a part of the transport system in earthworms. It is circulated by the mesothelial cilia or by contraction of muscles in the body wall. [6] Interestingly, the anti-tumor effect of ECF/macromolecules from earthworms such as *Eisenia foetida* and *Lumbricus rubellus* have been determined through *in vitro* and *in vivo* experimentations against colorectal and hepatocellular carcinoma respectively with encouraging results. [7, 8] Recent studies have demonstrated the optimum anti-tumor effect of ECF proteins of *Eudrilus eugeniae* (EE) on lung, colorectal, cervical and prostate cancer cell lines making it a species of interest for further research. However, hitherto, the anti-cancer potential of ECF or components of ECF on oral cancer cells has not been demonstrated until recently. [9–11] Furthermore, the mechanism of action and molecular structure of anti-cancer molecules from ECF of EE remain a largely unaddressed domain in oncology research.

As conventional drug discovery is a tedious and resource consuming process, contemporary computational techniques are being adopted to foster and accelerate the same with good precision [12]. The advent of *in-silico* simulations has enabled researchers to study the interactions between ligands/drug molecules/biomolecules and specific receptors through which they mediate their action. Molecular docking is one such technique to visualize the binding pocket of a receptor and identify crucial interactions exerted by the ligands/biomolecules. [13] Homology modeling is a technique employed to construct a 3D model of a protein or receptor of interest. Molecular dynamics is a simulation tool adopted to understand the duration of interaction between a ligand and receptor with a special emphasis on the influence of the ligand on the protein's structure and integrity. [14]

The present study was aimed at investigating the anti-cancer potential of a purified protein extracted from extracted from EE on oral cancer cells *in-vitro*. Initially we isolated the 18 kDa ECF protein (18-ECFP) from EE and delineated its molecular structure using proteomic techniques entwined with *in-silico* homology modeling. Further, the study ascertained the mechanism of cell death induced by 18-ECFP on oral cancer cells through *in-vitro* experiments. To conclude, the mechanistic analysis was confirmed by key interactions of the 18-ECFP with Caspase-3 and Capsase-8 demonstrated through *in-silico* techniques. To the best of our knowledge, this is a first of its kind study, specially designed to amalgamate advanced computational techniques with *in-vitro* experimental proofs in reconnoitering the anti-cancer potential of 18-ECFP from EE on oral cancer cells.

2. Materials & Methods

Ethical approval for the study was obtained from the RUAS (Ramaiah University of Applied Sciences), human and animal ethics committee. (No: FDS/EC/2014-16/PhD_03). The flow of experiments undertaken are illustrated in Flowchart 1.

2.1. Purification of Crude Extract of ECF of EE by SDS-PAGE

The crude extract of ECF of EE was partially purified to obtain a protein band in the range of 18–20 kDa. This band was loaded in lane 2 of the apparatus and subsequently was subjected to multiple SDS-PAGE trials which resulted in a specific 18 kDa band. The crude extract of ECF of EE was loaded in lane 1 and the standard protein marker was loaded in lane 4 for comparison. The separated 18 kDa protein band was precisely cut using a scalpel, vortexed and recovered from buffer solution for further proteomic analysis.

2.2. Characterization of 18-ECFP by MALDITOF & MS-MS Sequencing

Maldi-Tof analysis was performed to delineate the structure of the 18-ECFP. The 18 kDa band was isolated and subjected to MALDI-TOF & MS-MS sequencing using a Hybrid Instrument named MALDI-TOF/TOF MS Bruker Daltonics ULTRAFLEX III. The sample given in gel was trypsin digested and the peptides obtained were mixed with HCCA (α-Cyano-4-hydroxycinnamic acid) matrix in 1:1 ratio. The resulting 2 µl was spotted onto the MALDI plate. After air drying the sample, it was analyzed using the above-mentioned hybrid instrument. Maldi-Tof analysis was followed by MS/MS fragmentation of the respective peptides. Further analysis was executed with Flex Analysis Software for recording the MS & MS/MS Spectrum. The masses obtained post fragmentation of the peptides in MS & MS/MS were submitted for Mascot search against the given sequence for identification of the protein.

2.3. Proteomics Analysis of 18-ECFP by Nano-LCMS based Amino Acid Sequencing

The objective of proteomics is to characterize the entire range of proteins expressed in a cell or sample under defined laboratory conditions. The use of Nano-LCMS based amino acid sequencing for protein and peptide separations would result in greater efficiency in detailing the 18 kDa protein being analyzed. Q-TOF SYNAPT G2 Mass Spectrometer was employed for this purpose. The 18kDa gel band was used to extract the peptide with trypsin digestion. 10µl of the sample was injected on BEH C18 Ultra-High-Performance Liquid Chromatography (UHPLC) column for separation of peptides followed by analysis employing Q-TOF instrument for MS and MS/MS. The raw data was processed by Mass Lynx 4.1 WATERS, peptide editor software to get the complete integrated sequence of the sample. The individual peptides MS/MS spectra were matched with its database for amino acid sequence and protein identification on PLGS software, WATERS.

2.4. Homology Modeling of 18-ECFP

Homology modeling was employed to determine the 3-dimensional structure of the 18-ECFP. Due to unavailability of 3D structure in Protein Data Bank (PDB), the protein was modelled in Prime module via energy-based methods. (Schrodinger Release 2019-3: Prime, Schrodinger, LLC, New York, NY, 2019). Protein sequence was subjected to Basic Local Alignment *Search* Tool for protein (BLASTp) analysis in Uniprot, which revealed highest similarity with p-chain belonging to protein “giant hemoglobin” from species *“Glossoscolex paulistus”* (PDB ID: 4U8U) at threshold level 0.01 using the matrix Blosum-62. The protein was modelled using “giant haemoglobin” as template PDB by retaining the co-crystallized structures. The protein was aligned with template through Clustal W method and the protein structure energy was minimized using Schrodinger software. Finally, the modelled protein was validated by a Ramachandran Plot and subjected to Molecular Dynamics Simulation (MDS).

2.5. Molecular Dynamics Simulation of the Modelled Protein

To gain insights regarding the stability and dynamic behaviour of the modelled protein MDS was performed employing Desmond tool of Schrodinger software (2019-2) (Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2019). The system was prepared using TIP3P solvent model and set at normal pressure (1 bar) and temperature (300 K) ensemble for 50 ns with OPLS3e force field (Fig. 1A). Qualitative analysis predicting pressure, volume, temperature, total energy and potential energy were calculated using simulation quality analysis tool (Fig. 1B). Qualitative and quantitative assessments of modelled protein was studied using simulation event analysis tools. Stability of modelled protein was described in terms of Root Mean Square Deviation (RMSD) and Protein-Root Mean Square Fluctuations (P-RMSF).

RMSD is calculated based on changes in the structural deviations of the MP throughout the dynamics with reference structure whereas, RMSF is calculated based on the fluctuations of the amino acids with respect to initial structure in the protein chain throughout the simulations.

RMSD is calculated as:

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i'(t_x)) - r_i(t_{ref}))^2}$$

where N is the number of atoms in the atom selection; $tref$ is the reference time, (typically the first frame is used as the reference and it is regarded as time t=0); and r' is the position of the selected atoms in frame x after superimposing on the reference frame, where frame x is recorded at time tx.

The RMSF for residue i is calculated as:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < (r_i'(t)) - r_i(t_{ref}))^2 >}$$

where T is the trajectory time over which the RMSF is calculated,

$tref$ is the reference time

r_i is the position of residue i;

r' is the position of atoms in residue i after superposition on the reference

Protein Secondary Structure Elements (SSE) like alpha-helices and beta-strands are monitored throughout the simulation.

2.6. Docking of Modelled 18-ECFP with Pro-Apoptotic Receptors

This protein-protein docking between the modelled 18 kDa protein and caspase receptors would enable us to better understand the interaction between the modelled 18-ECFP with Caspase-3 and Caspase-8 receptors. The modelled protein prior to docking and the caspase receptors downloaded from PDB are shown in Figs. 2A-2F.

Protein-Protein Docking (PPD) and Molecular Mechanics-Generalized Born model and Solvent Accessibility (MM-GBSA) were performed after MD analysis. PPD was performed using Bio luminate suite (Schrodinger Release 2019-3: Bio Luminate, Schrödinger, LLC, New York, NY, 2019) with Caspase-3 (PDB ID: 6BDV) and Caspase-8 (PDB ID: 5JQE) to analyse the affinity of the modelled protein towards selected anti-cancer targets. The above-mentioned analysis have been performed for the first time for a component of the ECF with anti-cancer effect.

2.7. MM-GBSA Analysis of Docked Poses of Modelled 18-ECFP with Caspase Receptors

The top 5 poses exhibiting highest PIPER score were subjected to MM-GBSA energy calculations using Prime tool of Schrodinger. MM-GBSA calculations were performed for the assessment of binding free energies (ΔG_{bind}) for the best docked compounds employing the OPLS3e force field. Prime/MM-GBSA (Schrodinger Release 2019-3: Prime, Schrödinger, LLC, New York, NY, 2019) method using the below mentioned equation.

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{SOL} + \Delta G_{SA}$$

where ΔE_{MM} is the difference in energy between complex structure and sum of energies of ligand and unliganded protein using OPLS force field. ΔG_{SOL} is the difference in GBSA solvation energy of protein-inhibitor complex and sum of solvation energies for unbound protein and ligand. ΔG_{SA} is the difference in surface area energies for complex and sum of surface area energies for unbound protein and ligand (Lyne et al., 2006).

2.8. Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

A semi quantitative RT-PCR was performed to determine the gene expression levels of Caspase-3 and Caspase-8 in SCC-9 cells induced by 18-ECFP. Standardization was performed using β-actin gene in SCC-9 cells (Fig. 3A). SCC-9 cells were cultured in DMEM supplemented with HEPES, using 10% heat inactivated foetal bovine serum (FBS), 100 units/ml penicillin G, and 100µg/ml streptomycin at 37°C, 5% CO_2 incubator. Two concentrations (10 µg and 20 µg) of the test sample were employed for testing. Primers are designed to anneal two opposite strands of specific nucleic acid template targets, they are usually between 15–40 bases long. Primers for the current experiment was designed and synthesized at Eurofins Genomics, India as shown in Table 1.

Table 1
Primer Details, Primers Synthesized at Eurofins Genomics, India

Gene	Primer pair	Sequence	Tm	Product size (bp)
β – Actin	FP	TCCTCCTGAGCGCAAGTACTCT	62.1	153
	RP	GCTCAGTAACAGTCCGCCTAGAA	62.4	
Bcl-2	FP	CTGGTGGACAACATCGCTCTG	61.8	228
	RP	GGTCTGCTGACCTCACTTGTG	61.8	
Bax	FP	GACACCTGAGCTGACCTTGG	61.4	310
	RP	GAGGAAGTCCAGTGTCCAGC	61.4	
Caspase-3	FP	ACATGGCGTGTCATAA AATACC	51.1	120
	RP	CACAAAGCGACTGGATGAAC	51.8	
Caspase-8	FP	CATCCAGTCACTTTGCCAGA	51.8	128
	RP	GCATCTGTTTCCCCATGTTT	49.7	

Sample Preparation and RNA Isolation: SCC-9 cells were washed twice with PBS and to the adherent cells 2 ml of TRIzol (per T25 flask) was added and transferred to the tube and vortexed. 0.2 ml of chloroform was added per 1ml of TRIzol. The colourless upper aqueous phase was transferred to a new clean tube. 0.5 ml of isopropanol was added per 1 ml of TRIzol used, the sample was mixed gently by inverting the sample 5 times and incubated at room temperature for 5 minutes followed by centrifugation. The supernatant was discarded and the RNA pellet was washed by adding 1 ml of 70% ethanol followed by centrifugation. The supernatant was discarded by inverting the tube on a clean tissue paper. Later, the pellet was dried by incubating in a dry bath for 5 minutes at 55°C. The pellet was then resuspended in 25 µl of DEPC treated water

The PCR mixture (final volume of 20 µL) contained 1 µL of cDNA, 10 µL of Red Taq Master Mix 2x (Amplicon) and 1µM of each complementary primer specific for Caspase-3, Caspase-8, Bcl-2, Bax and β-Actin (internal control) sequence. The samples were denatured at 94°C for 5 minutes, and amplified using

30 cycles at 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute. Annealing temperature for Caspase-3 and Caspase-8 was set to 58°C for 30 seconds, for Bcl-2 and Bax it was set to 56°C and for β-Actin the annealing was set at 55°C for 30 seconds followed by a final elongation at 72°C for 10 minutes. Optimal number of cycles were selected for amplification of Capsase-3, Caspase-8, Bcl-2 and Bax so that amplifications in the exponential range are obtained. 10 µl of the final amplification end product were made to run on 2% agarose gel stained with ethidium bromide. Quantification was performed by measuring the optical density of the bands using Image J software.

2.9. Quantitative Real-Time Polymerase Chain Reaction (Q-PCR)

A real time quantitative Q-PCR was performed to determine the gene expression levels of Caspase-3 and Caspase-8 in SCC-9 cells induced by 18-ECFP. For each of the target genes, PCR conditions *viz*, Tm, amplicon specificity and size were optimized using in-house established and validated methods. The melt peak profile is presented in Fig. 3B and melt temperature of target genes is shown in Table 2. The quantification of gene expression was performed by quantitative real-time (Q-PCR) by using the CFX96 real time PCR machine and SYBR qPCR mix. The following conditions were used: 40 cycles at 94°C for 15 sec, 54°C for 20 sec, and 72°C for 30 seconds. To assess the quantity of isolated RNA and the efficacy of cDNA synthesis, target cDNAs were normalized to the endogenous mRNA levels of the housekeeping gene GAPDH. PCR amplifications were performed at least 3 times for each sample.

Table 2
Melt Temperature of target genes

Target Gene	Treatment	Melt Temp (°C)
Caspase-3	Control	79.5
Caspase-3	10 µg/ml	79.5
Caspase-3	20 µg/ml	79.5
Caspase-8	Control	83
Caspase-8	10 µg/ml	83
Caspase-8	20 µg/ml	83
GAPDH	Control	76
GAPDH	10 µg/ml	76
GAPDH	20 µg/ml	76

3. Results

3.1. Purification of Crude Extract of ECF of EE by SDS-PAGE

The SDS-PAGE analysis facilitated the isolation of 18-ECFP. The crude extract of ECF of EE in lane 1 showed moderately thick bands without clear distinction between protein ranges. As lane 2 contained the already purified 18–20 kDa protein sample, further separation revealed finer and more specific bands. (Fig. 4A) The 18 kDa band from lane 2 was cut using a scalpel, the protein was recovered from solution and stored at -80°C.

3.2. Characterization of 18-ECFP by MALDITOF & MS-MS Sequencing

Protein Mass Fingerprint (PMF) of the trypsin digested 18-ECFP was obtained through Malditof-MS analysis as shown in Fig. 4B. After fragmentation of the parent compound from MS the molecular weight of parent compound was confirmed to be in the range of 18 kDa. The parent compound was again fragmented in MS2, shown in Figs. 4C-4E. A Mascot search was applied to the given sequence for significant identification of the protein for the masses that were obtained by fragmentation of the peptides in MS & MS/MS. The details of the Mascot search results are shown below, the number 96 represents the highest matching score.

The MALDI-TOF – MS/MS analysis of the sample 18-ECFP showed peptide match with Extracellular globin OS = *Glossoscolex paulistus* (details shown below)

Database : Lumbricina 20180918 (4677 sequences; 992615 residues)

Timestamp : 18 Sep 2018 at 10:40:31 GMT

Top Score :96 for A0A0P4VL51_9ANNE, Extracellular globin OS = Glossoscolex Paulistus OX = 1046353 GN = HgBp PE = 3 SV = 1

Mascot Score Histogram and match – shown in Fig. 5A

Protein score is -10*Log(P), where P is the probability that the observed match is a random event. Protein scores greater than 49 are significant ($p < 0.05$). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

This data was insufficient as several amino acid details were missing, hence a complete amino acid sequencing preceded by a Nano LC-MS was performed to derive information regarding all the amino acids present in the protein.

[illegible]o-LCMS Based Amino Acid Sequencing

The UHPLC revealed peptide information regarding the 18-ECFP, shown in Fig. 5B. The UHPLC was followed by MS-MS analysis by employing the Q-TOF instrument and it revealed the spectrum illustrated in Fig. 5C. The obtained MS/MS spectra of the peptides revealed significant information on the AA sequence of 18-ECFP, this peptide information was matched against existing databases to confirm the identified sequence. Through the experiments conducted, a high-resolution MS/MS spectrum of the peptides of the 18-ECFP were obtained which were further subjected to AA sequencing.

This was the first attempt to determine the AA sequence of a protein from the ECF of EE. The AA sequencing enabled the identification of a total of 158 amino acids in the protein with precise sequence, as shown in Fig. 6.

Details from PLGA, WATERS:

Associated Spectrum Datafile: 12022019_Chromous_D_20

Mass Window = 0.50 amu, Threshold = 10000

Average Mass = 18401.0127, Monoisotopic Mass = 18389.3678

Residues: 1-158

N-Terminus = H, C-Terminus = OH

Fragment ions: Monoisotopic/Average (2) m/z ratios with 1 positive charge(s).

The complete mass list was found to be as shown in Table 3. The peptides of the sample generated were sequenced by QTOF MSMS analysis, represented below:

MKPQIVVFIVLVSYVSGQCSILESLKVKNQWADAFGKGNRRVEFGLRLWNSFFQDHPETRDLFKRVRGDNAYSPEFKAHAQRVLSGFDMTVSLLDDPDTFAAQIAHLKKQHD

A total of 158 amino acids in accurate sequence was identified through the experiments conducted.

Table 3
Complete Mass List of the 18-ECFP Analyzed

Sl No	Start-End	Mass (Da)	Sequence
1.	1–28	3180.5537	MKPQIVVFIVLVSYVSGQCSILESLKVK
2.	29–37	1036.6053	NQWADAFGK
3.	38–41	502.3095	GNRR
4.	42–47	720.4978	VEFGLR
5.	48–64	1067.5116	LWNSFFQDHPETRDLFK
6.	61–65	678.4706	DLFKR
7.	66–77	1382.7778	VRGDNAYSPEFK
8.	78–108	3383.702	AHAQRVLSGFDMTVSLLDDPDTFAAQIAHLK
9.	106–118	1642.1356	HLKKQHDPRNLKK
10.	110–117	1007.582	QHDPRNLK
11.	118–125	1190.7621	KEYFDWFR
12.	125–138	1668.0195	RNHLLEILPEYIGT
13.	128–144	1964.0309	LLEILPEYIGTKLDFEA
14.	140–158	2218.0698	LDFEAWTHCFNHISSGISP

3.4. Homology Modeling of 18-ECFP

Based on the Nano-LCMS - AA sequencing analysis, a homology model was built, which is represented in Figs. 7A-7C. This model was constructed in order to dock this protein of interest with apoptotic receptors Caspase-3 and Caspase-8.

The BLASTp analysis in Uniprot revealed that p chain of 4U8U, crystallographic structure of the giant hemoglobin from "*Glossoscolex paulistus*", with resolution 3.2 A° was the most suitable as template as it exhibited 51% identities, 61% positives and 12% gaps with the sequence. The sequence was aligned through Clustal W method and the protein structure energy was prepared using Schrodinger software. The protein was modelled through energy-based method and validated by a Ramachandran Plot (Fig. 7D). The plot revealed that ~ 95% of amino acids were found to be in most favoured regions (Fig. 7E).

3.5. Molecular Dynamics Simulation of the Modelled Protein

The MP exhibited stability after 10 ns and no major fluctuations were observed in RMSD values throughout the simulation period, shown in Figs. 8A & 8B. P-

tions in initial 20 residues as expected due to terminal residues. The residues involved in formation of secondary

structure exhibited less fluctuations. More than half of residues (53.45%) of the modelled protein were found to be involved in α-helix formation.

Secondary Structure Elements (SSE): Alpha-helical and beta-strand regions are highlighted in red and blue backgrounds, shown in Fig. 9A. These regions are defined by helices or strands that persist over 70% of the entire simulation. The figure summarizes the SSE composition for each trajectory frame over the course of the simulation.

Figure 9B reports SSE distribution by residue index throughout the protein structure monitoring each residue and its SSE assignment over time. This is the first report of MDS that has been performed for a specific component of any earthworm extract. The simulated protein is shown in its *in-silico* matrix in Fig. 9C & 9D.

3.6. Docking of Modelled 18-ECFP with Pro-Apoptotic Receptors

The 18-ECFP revealed significant interactions with Caspase-3 and Caspase-8 receptors indicating apoptotic activation in pathway-based lysis of SCC-9 cells. Docking analysis with 6BDV (Caspase-3 receptor) and 5JQE (Caspase-8 receptor) revealed 30 PIPER poses (resembles the docking poses for every pose between the 2 proteins) with scores ranging from − 82.549 to -606.115 and 364.946 to -278.548 respectively, which are shown in Table 4 and Table 5 (Top 5 pose scores are bolded) accordingly.

Table 4
Piper Pose Scores for PPD between 18-ECFP and 6BDV Receptor with their Corresponding Poses (Top 5 poses have been highlighted)

Sl No	Title	PIPER Pose Energy	PIPER Pose Score	PIPER Model Number
1.	prot-prot-docking_6bdv_pose_1	-791.96	-248.028	414
2.	prot-prot-docking_6bdv_pose_2	-792.769	-301.145	402
3.	**prot-prot-docking_6bdv_pose_3**	-791.133	**-407.64**	424
4.	prot-prot-docking_6bdv_pose_4	-842.955	-246.976	134
5.	prot-prot-docking_6bdv_pose_5	-858.03	-279.015	95
6.	**prot-prot-docking_6bdv_pose_6**	-778.883	**-351.607**	567
7.	prot-prot-docking_6bdv_pose_7	-899.338	-230.557	39
8.	prot-prot-docking_6bdv_pose_8	-872.855	-175.24	69
9.	prot-prot-docking_6bdv_pose_9	-767.834	-179.777	704
10.	prot-prot-docking_6bdv_pose_10	-769.136	-275.384	693
11.	prot-prot-docking_6bdv_pose_11	-754.23	-140.074	901
12.	prot-prot-docking_6bdv_pose_12	-752.037	-236.293	935
13.	prot-prot-docking_6bdv_pose_13	-757.94	-234.476	832
14.	**prot-prot-docking_6bdv_pose_14**	-753.036	**-323.082**	920
15.	prot-prot-docking_6bdv_pose_15	-756.14	-82.549	863
16.	prot-prot-docking_6bdv_pose_16	-769.865	-185.714	684
17.	prot-prot-docking_6bdv_pose_17	-772.183	-175.801	652
18.	**prot-prot-docking_6bdv_pose_18**	-792.023	**-606.115**	412
19.	prot-prot-docking_6bdv_pose_19	-792.524	-248.831	405
20.	prot-prot-docking_6bdv_pose_20	-841.251	-242.847	145
21.	prot-prot-docking_6bdv_pose_21	-805.387	-256.097	315
22.	prot-prot-docking_6bdv_pose_22	-788.86	-187.8	453
23.	prot-prot-docking_6bdv_pose_23	-784.293	-241.494	500
24.	prot-prot-docking_6bdv_pose_24	-827.826	-303.484	198
25.	prot-prot-docking_6bdv_pose_25	-775.196	-142.662	614
26.	prot-prot-docking_6bdv_pose_26	-748.42	-211.688	992
27.	prot-prot-docking_6bdv_pose_27	-761.295	-208.453	793
28.	**prot-prot-docking_6bdv_pose_28**	-756.535	**-305.217**	856
29.	prot-prot-docking_6bdv_pose_29	-752.569	-96.623	929
30.	prot-prot-docking_6bdv_pose_30	-806.753	-230.283	302

Table 5
Piper Pose Scores for PPD between 18-ECFP and 5JQE Receptor with their Corresponding Poses (Top 5 poses have been highlighted)

Sl No	Title	PIPER Pose Energy	PIPER Pose Score	PIPER Model Number
1.	prot-prot-docking_5jqe_pose_1	-748.143	-120.059	333
2.	prot-prot-docking_5jqe_pose_2	-765.602	-167.396	235
3.	prot-prot-docking_5jqe_pose_3	-843.683	-20.095	69
4.	**prot-prot-docking_5jqe_pose_4**	-791.016	**-230.801**	137
5.	prot-prot-docking_5jqe_pose_5	-705.623	-209.504	748
6.	**prot-prot-docking_5jqe_pose_6**	-741.73	**-219.31**	368
7.	prot-prot-docking_5jqe_pose_7	-752.327	-185.754	307
8.	**prot-prot-docking_5jqe_pose_8**	-754.907	**-278.548**	283
9.	prot-prot-docking_5jqe_pose_9	-721.677	-80.887	533
10.	prot-prot-docking_5jqe_pose_10	-702.7	-154.792	795
11.	prot-prot-docking_5jqe_pose_11	-773.503	-45.709	196
12.	**prot-prot-docking_5jqe_pose_12**	-775.472	**-221.794**	185
13.	**prot-prot-docking_5jqe_pose_13**	-764.554	-104.85	243
14.	prot-prot-docking_5jqe_pose_14	-742.424	-34.516	365
15.	prot-prot-docking_5jqe_pose_15	-702.52	-120.967	797
16.	prot-prot-docking_5jqe_pose_16	-720.017	-60.7	550
17.	prot-prot-docking_5jqe_pose_17	-709.082	-160.12	697
18.	prot-prot-docking_5jqe_pose_18	-701.974	-154.33	814
19.	prot-prot-docking_5jqe_pose_19	-790.765	50.436	138
20.	prot-prot-docking_5jqe_pose_20	-731.976	-95.038	442
21.	prot-prot-docking_5jqe_pose_21	-695.084	-154.789	942
22.	prot-prot-docking_5jqe_pose_22	-740.811	-90.606	379
23.	prot-prot-docking_5jqe_pose_23	-798.908	-9.065	118
24.	prot-prot-docking_5jqe_pose_24	-772.483	364.946	199
25.	prot-prot-docking_5jqe_pose_25	-706.6	-87.187	727
26.	prot-prot-docking_5jqe_pose_26	-696.166	-129.746	923
27.	prot-prot-docking_5jqe_pose_27	-707.576	-101.457	711
28.	**prot-prot-docking_5jqe_pose_28**	-702.243	**-272.288**	802
29.	prot-prot-docking_5jqe_pose_29	-702.488	-98.495	799
30.	prot-prot-docking_5jqe_pose_30	-693.662	-87.721	965

PPD of 18 kDa protein with 6BDV – Caspase-3 receptor at pose_18 showed the highest PIPER score of -606.115. PPD of 18 kDa protein with 5JQE – Caspase-8 receptor at pose_8 displayed the highest PIPER score of -278.548. The Figs. 10A-10E and 11A-11E represent the 5 best poses of the PPD between 18-ECFP with Caspase-3 and Caspase-8 receptors. The modelled protein is depicted in cyan, the caspase receptors are shown in pink and tri-colour (red, blue, white). The colours red, blue and white represent hydrogen acceptors, donors and neutral atoms in Caspase-3 and Caspase – 8 receptors.

3.7. MM-GBSA Analysis of Docked Poses of Modelled 18-ECFP with Caspase Receptors

The G_{bind} for the top 5 poses obtained in PPD between 18- ECFP with 6BDV – Caspase 3 receptor and 5JQE – Caspase 8 receptor is presented in Table 6. In the current study the best binding score was exhibited at pose_28 between the 18 kDa protein and 6BDV – Caspase 3 receptor with a binding energy of -93.73 kcal/mol. Whereas a high binding energy of -103.21 kcal/mol was witnessed at pose_28 between the 18 kDa protein and 5JQE – Caspase 8 receptor.

Table 6
The Prime Energy and Bind Score Calculation for the Top 5 Poses Obtained in PPD

Sl No	PPD POSE	ΔBind energy	Δ Coloumb	ΔG Covalent	ΔG H bond	ΔG-vdW
MM-GBSA 6bdv – Caspase 3 Receptor						
1.	pose_18	**-59.38**	167.432	30.97	3.7	-116.11
2.	pose_3	**-69.91**	5.81	14.80	-6.04	-72.48
3.	pose_6	**-72.07**	85.38	24.50	-7.48	-88.20
4.	pose_14	**-41.94**	47.58	38.66	-2.95	-87.55
5.	pose_28	**-93.73**	-11.60	25	-4.47	-97.81
MM-GBSA 5jqe – Caspase 8 Receptor						
1.	pose_8	**-60.62**	653.81	9.83	-6.15	-154.04
2.	pose_28	**-103.21**	80.28	26.01	-6.19	-100.93
3.	pose_4	**1773.77**	424.85	974.81	-8.56	792.83
4.	pose_12	**-69.89**	47.87	26.51	-3.38	-91.59
5.	pose_6	**-23.8**	304.4482	-2.48767	-0.77	-49.38

3.8. Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

A semi quantitative RT-PCR was performed using Techno Prime system to determine the levels of Caspase-3, Caspase-8, Bcl-2, Bax and β-Actin mRNA expressions. The cDNA was synthesized from 2 µg of RNA using the Verso cDNA synthesis kit (Thermo Fischer Scientific) with oligo dT primer according to the manufacturer's instructions. The reaction volume was set to 20 µl and cDNA synthesis was executed at 42°C for 60 min, followed by RT inactivation at 85°C for 5 min.

In the present study the gene expression of Caspase-3, Caspase-8, Bcl-2 and Bax was assessed on SCC-9 cells after incubating with 18-ECFP. The analysis of amplicons revealed the following (Fig. 12A): Caspase-3 was upregulated by 1.47-fold and 2.05-fold in SCC-9 cells at a sample test concentration of 10 and 20 µg/ml respectively as shown in Table 7 & Fig. 12B. Caspase-8 was upregulated by 1.76-fold and 2.07-fold in SCC-9 cells at a sample test concentration of 10 and 20 µg/ml respectively as shown in Table 8 & Fig. 12C. Through the results obtained it is evident that the 18-ECFP has upregulated proapoptotic genes in SCC-9 cells.

Table 7
Relative Expression of Caspase-3 in Treated and Untreated SCC-9 Cells

Sample Conc.	Band Intensity of PCR Amplicon of Genes		Normalized Caspase-3 expression	Relative Gene Expression
	β-Actin	Caspase-3		
Control	26886.54	19600.09	0.73	1.00
10 µg/ml	19340.17	20698.13	1.07	1.47
20 µg/ml	19036.33	28493.22	1.50	2.05

Table 8
Relative Expression of Caspase-8 in Treated and Untreated SCC-9 Cells

Sample Conc.	Band Intensity of PCR Amplicon of Genes		Normalized Caspase-8 expression	Relative Gene Expression
	β-Actin	Caspase-8		
Control	26886.54	28273.38	1.05	1.00
10 µg/ml	19340.17	35720.45	1.85	1.76
20 µg/ml	19036.33	41472.45	2.18	2.07

Bcl-2 was not upregulated by a big margin, a 1.05-fold and 1.37-fold in SCC-9 cells at a sample test concentration of 10 and 20 µg/ml respectively was observed. (Table 9, Fig. 13A and 13B) This was in accordance with the pro-apoptotic effect of the 18-ECFP as Bcl-2 is an anti-apoptotic protein and its upregulation is not favoured. Bax gene did not show any upregulation.

Table 9
Relative Expression of Bcl-2 in Treated and Untreated SCC-9 Cells

Samples	Band Intensity of PCR Amplicon of Genes		Normalized Bcl-2 expression	Relative Gene Expression
	Beta-Actin	Bcl-2		
Control	55101.39	57752.768	1.05	1.00
10 µg/ml	71532.48	60450.91	1.10	1.05
20 µg/ml	72299.32	79146.839	1.44	1.37

3.10. Quantitative Real-Time Polymerase Chain Reaction (Q-PCR)

The upregulation of Caspase-3 expression at 10 µg/ml and 20 µg/ml was found to be 1.414 and 2.603 folds whereas, the Caspase-8 expression at 10 µg/ml and 20 µg/ml was observed to be 0.500 and 1.866 folds. Overall, Caspase-3 and Caspase-8 expression was found to be concentration dependent, as shown in Table 10 & Fig. 14.

Table 10
Q-PCR Analysis of Caspase-3 and Caspase-8 in SCC-9 Cell Line

Target Gene	Treatment	Gene Expression	Mean Cq or Ct values
Caspase-3	Control	1.000	34.33
Caspase-3	10 µg/ml	1.414	31.25
Caspase-3	20 µg/ml	2.603	31.4
Caspase-8	Control	1.000	29.29
Caspase-8	10 µg/ml	0.500	27.71
Caspase-8	20 µg/ml	1.866	26.84
GADPH	Control	N/A	23.9
GADPH	10 µg/ml	N/A	21.32
GADPH	20 µg/ml	N/A	22.35

4. Discussion

Drug discovery is an arduous and expensive process, with long term preclinical and clinical phases. Currently, the drug industry is witnessing a paradigm shift towards a pathology-based bioinformatics approach. The advent of artificial intelligence and machine learning programmed softwares provide a better understanding of molecular interactions of drugs with receptors. Through the application of computational biology, the role of drug molecules in the pathophysiology of diseases can be virtually predicted with precision. [15]

The development of *in-silico* simulation has now made it possible to infer detailed information regarding pathways of regulatory genes, mechanism of action of potential drug candidates and binding energy calculations between a molecule and its receptor. It is likely that the development of these technologies will dramatically change, the drug discovery process in the next few decades. [16, 17]

The research work undertaken in this study describes the 3D homology model construction of 18-ECFP and its molecular dynamics simulation. The current study has employed computational biology to examine the interaction of the modelled protein with cancer cell death receptors alongside binding energy calculations. The gene expression induced by 18-ECFP in oral cancer cell line SCC-9 *in vitro* has also been determined by wet lab techniques providing an insight into the mechanism of cell death.

Proteomic procedures were undertaken to determine the type and sequence of amino acids present in the 18-ECFP for modeling by energy-based methods. These findings enabled the construction of the homology model of 18-ECFP that is required for *in-silico* evaluation (molecular docking and binding energy calculations) with cancer cell death receptors Caspase-3 and Capsase-8.

The isolation of 18-ECFP was initiated by SDS-PAGE and progressed by MALDITOF & MS-MS Sequencing. The technique of PMF involves digestion of a target protein with trypsin and obtaining a mass spectrum of the peptides in the mixture. [18] The m/z values obtained were crossmatched with the known protein databases for a parallel match. This method exclusively relies on the peptide map obtained in the mass spectrum and not on the AA sequence. Proteins with translational modifications can also be identified with a PMF analysis, however increased post translational changes may interfere with the protein identification. [19] Tandem Mass spectrometry (TMS) or Mass spectrometry (MS) is an advanced technique for protein analysis and is an indispensable tool for proteomics research. [20] For any MS based experiment, the instrument type, the fragmentation strategy and method of analysis should be considered. The MS functions on the principle of charge to mass ratio (m/z), the MS measures the (m/z) of ions in gas phase. The MS consists of a source of ions that convert the protein analyte into gas phase ions. A mass analyzer separates the ionized analytes based on their (m/z) ratio. The detector records the number of

Protein analysis by MS was revolutionized with the advent of electrospray ionization (ESI) and matrix-assisted laser desorption/ ionization (MALDI), these soft ionization techniques are capable of ionizing peptides or proteins efficiently, the same has been utilized in the current study. Use of two simultaneous mass analysers increases the detection specificity and helps in detecting target proteins with precision. [22, 23] In this scenario the first analyzer depends on the data of the other analyzer, hence it's also referred to as data dependent.

A study by Pisamai et al., 2018, unveiled the unique PMFs of novel candidate protein markers of canine oral tumors by performing Malditof and LCMS analysis. The authors identified several peptide fragments that play a crucial role in tumorigenesis. The authors concluded by stating that this data might help veterinarians choose drugs of choice and treatment plan. [24] In the present study Malditof-MS was employed to determine the PMF of the 18-ECFP, this was implemented in order to study the peptides present. This is the first report of Malditof-MS analysis of ECF from EE. The Malditof – MS/MS analysis followed by a Mascot search of the sample 18-ECFP showed peptide match with Extracellular globin OS = *Glossoscolex Paulistus* (Appendix – 1). However, several amino acid details were missing hence a Nano-LCMS bases amino acid sequencing had to be performed to retrieve complete sequence and number of amino acids present in the 18 kDa protein.

Hybrid platforms like LC-MS/ MS and high-resolution LCMS are popular analytical methods for productive research. In LC-MS-based proteomics, the complex mixture of proteins is first enzymatically cleaved resulting in formation of peptide products which are then analyzed by a mass spectrometer. A small percentage of peptides present in the sample is usually considered for identification. The 10 most abundant peaks are selected from the first MS spectra (MS1) step for fragmentation in the second MS analysis (MS2).The data obtained is simplified by removing redundant information from isotopic peaks through a process called deisotoping. [25]

The present study has utilized the Q-TOF SYNAPT G2 Mass Spectrometer. The equipment provides the most complete characterization of complex mixtures and molecules with unique levels of MS performance, industry leading informatics and unparalleled platform versatility with superior separation with high-efficiency Ultra Performance Liquid Chromatography (UPLC)-MS/MS performance. The current study preferred UPLC over HPLC for protein analysis. The pressure difference applied differentiates between the High-Performance Thin Layer Chromatography (HPLC) and UPLC procedures. The development of < 2 microns LC columns have drastically increased the efficiency of the LC procedure. [26, 27]

UPLC employs robust high efficiency pumps perform at ultra-high pressures hence the name. Due to the capability of high-speed usage, dwell volume, detector flow cell and detector scanning rate the UPLC's were designed using significantly smaller diameter tubing. The UPLC sensors have increased scanning rate efficiency which ensures that the entire protein peak could be detected with no missing data. [28] Conventional HPLC instruments employ scanning rates of 20 Hz only whereas UPLC detectors use high rates at 160 Hz. [29] The use of a UPLC instrument like the one employed in the current study ensured production of reliable results.

The performance of MS technology in the domain of oncology provides a greater understanding of pathology in terms of new protein markers and possible drug targets. [30] A study by Fiolka et al., 2019, determined the amino acid sequence of a protein carbohydrate complex from the coelomic fluid of earthworm *Dendrobaena veneta.* The authors performed LC-MS screening-based AA-sequencing to delineate the fundamental composition of the complex. The authors concluded that it can be used for the treatment of skin and mucous membrane candidiasis in the future due to its anti-fungal activity. [31] The current study performed AA-sequencing for the first time for a coelomic fluid protein of EE. The 18-ECFP revealed a total of 158 amino acids making this the first report in literature. (Fig. 6 and Table 3)

The knowledge of the type, number and sequence of amino acids present enabled a homology model construction. Homology modeling is a tool for drug discovery that aids in construction of a 3D model of the molecule of interest. Using simulation tools, the constructed model can be docked with known pro-apoptotic cancer receptors, since anti-cancer drugs act through these receptors in causing cancer cell death. Identifying the structure of these biomolecules could be an initial step towards drug development and future anti-cancer research expectations could be promising.

A recent study conducted by Abdelmonsef AH, 2019, built a valid three-dimensional (3D) model of Rab39a. The amino acid sequence of Rab39a (ID: Q14964) was retrieved from Uniprot database. The Rab39a/DENND5B interactions were examined by molecular protein–protein docking. The authors concluded that Rab39a has emerged as a therapeutic target for drug development towards lung cancer. [32]

Compared to the study stated above, the current study constructed a homology model not by merely downloading the sequence from a protein database, instead the complete sequence of the protein of interest was systematically determined in the laboratory by nano-LCMS based amino acid sequencing and then the homology model was built using Schrodinger software.

Schrodinger is a state-of-the-art computational biology technology that encompasses homology modeling and protein sequence analysis tools that include advanced loop predictions, chimeric model building, annotation capabilities and interactive protein structure quality analysis. [33]

The current study has used Schrodinger simulation software to build the previously unknown homology model of the 18-ECFP. The structure of ECF anti-cancer protein was previously neither available nor modelled. To the best of our knowledge this is the first report of a homology model of any earthworm extract till date. (Fig. 7A-7D) This highlights the novelty of the present study.

As noted from literature review, only few studies have employed MDS to demonstrate the dynamics of an anti-cancer protein in question. The MD simulation was carried out using Desmond tool of Schrodinger software. A study by Sreenivasan et al., 2014, performed MDS for a proposed anti-cancer protein Nek6. The overall structural fluctuation of Nek6 protein was evaluated by analysing the RMSDs of backbone atom versus simulation time. The authors stated that the trajectories produced a stable protein that could be carried forward for molecular docking. [34] A study by Nagpal et al., 2017, evaluated the mortalin-p53 complex formation by MDS. Mortalin with sequence ranging from 50–384 showed an average standard deviation of 0.303 in the RMSD profile of the

modelled protein structure. The authors stated that the simulated protein structure had not deviated much in comparison to its initial structure and is relatively stable. [35]

The predicted model of 18-ECFP in the present study was proved to be stable and reliable after MDS evaluation in terms of spatial and energy boundaries. This is the first report of MDS that has been performed for a specific component of any earthworm extract. (Fig. 8 and Fig. 9) This is an added novelty of this study undertaken. The 18 kDa model was subjected to molecular docking studies. The modelled protein was docked with pro-apoptotic cancer receptors to study the binding affinity. Results of the current study exhibited satisfactory binding affinity with pro-apoptotic genes Caspase-3 and Caspase-8 responsible for cancer cell apoptosis.

It is crucial to predict protein-ligand or protein-protein interactions with accurate assessment of their binding energies. The BioLuminate programme of Schrodinger offers a comprehensive strategy for protein-protein docking experiments. [36] Built on a compact groundwork of all-inclusive protein modeling tools, BioLuminate offers access to supplementary advanced tools for protein engineering and analysis of protein-protein interactions. [37] Schrodinger's BioLuminate is the first comprehensive user interface tool designed with significant user input, to precisely address the key requests related with the molecular design of biologic samples. [38]

A recent study by Tahlan et al., 2019, screened benzimidazole compounds for their anti-cancer property. The docking study of data sets were carried out by Schrodinger-Maestro v11.5 using CDK-8 (PDB code: 5FGK) and ER-alpha (PDB code: 3ERT). Compound 12 exhibited the best docking score of − 8.907 with CDK-8 receptor. [39] Another recent study by Yadav et al., 2020, evaluated PPD of tubulin protein with paclitaxel, etoposide and topotecan by molecular docking using Schrodinger software. The authors concluded that etoposide is the best drug for tubulin with a docking score of – 4.916. [40]

The above-mentioned studies performed docking of molecule of interest whose crystal structure was already available in the PDB. The findings of the current study are unique as compared to the above-mentioned studies as the homology model of the protein of interest was first constructed and then docked with Caspase – 3 and Caspase – 8 receptors. (Fig. 10 and Fig. 11) Over 100 docking poses were witnessed for both Caspase – 3 and Caspase – 8 receptors with the homology model of the 18 kDa protein. The top 5 poses with the highest PIPER score were considered for energy calculations.

PIPER is a state-of-the-art protein-protein docking program based on a well-constructed multi-staged tactic and progressive numerical system that reliably generates precise structures of protein-protein complexes. Based on well-validated docking code adopted from the Vajda lab at Boston University, PIPER has a confirmed track record as an outstanding predictor of protein-protein complexes. These results have also been judged by previous CAPRI (Critical Assessment of Prediction of Interactions) blind experiments. [41]

The binding energies between the 18 kDa protein and caspase receptors were calculated by the MM-GBSA assay. The Prime MM-GBSA approach is employed to envisage the free energy of binding for a receptor with a set of ligands or proteins. Prime MM-GBSA generates a sizeable proportion of energy properties. The properties generated report interaction energies for the ligand, receptor, and protein-complex structures as well. The tool also generates energy differences pertaining to strain and binding. [42] In the current study the ΔG_{bind} for pose_28 between 18 kDa protein of ECF of EE with 6bdv – Caspase-3 receptor exhibited a top bind score of -93.73 kcal/mol. The ΔG_{bind} for pose_28 between 18 kDa protein of ECF of EE with 5jqe – Caspase-8 receptor exhibited a top bind score of -103.21 kcal/mol. Satisfactory binding between the modelled protein and caspase death receptors were witnessed through the simulations conducted. As the MM-GBSA binding energies are approximate free energies of binding, a more negative value indicates a stronger binding between two molecules. (Table 6)

A recent study by Lokhande et al., 2019, successfully docked Deguelin (CID: 107935) with cyclin D1 and cyclin E receptors. The authors employed the MM-GBSA to evaluate the ΔG_{bind} between the deguelin and respective cyclin D1 as well as cyclin E receptor. The authors reported a ΔG_{bind} of − 10.2 kcal/mol and − 8.8 kcal/mol for deguelin with cyclin D1 and cyclin E receptors respectively. [43]

Another recent study by Suganya et al., 2019, docked Proanthocyanidin (PAC) a promising anti-cancer compound with BCL-XL, CDK2 and were compared with 5-FU receptor. The authors employed the MM-GBSA and concluded that PAC exhibited better binding affinity of − 5.23, − 5.17 and − 4.43, − 4.47 kcal/mol against BCL-XL, CDK2 when compared to 5-FU. [44] This is the first of a kind novel report highlighting the interaction of an anti-cancer protein from an earthworm source and human caspase receptors with binding energy calculations.

After obtaining the *in-silico* results, the 18-ECFP was evaluated on oral cancer cell line SCC-9 *in vitro* to determine gene expression of apoptotic genes. Wet lab studies such as RT-PCR, and Q-PCR were undertaken to achieve this objective.

Through PCR reactions, target nuclei acid sequences can be amplified with the use of a DNA polymerase, primers and nucleotides. A nucleic acid sequence can be used as a template for a PCR reaction. The source for the nuclei acid could be DNA, RNA, or cDNA. Short nucleic acids synthesized *in vitro* are called primers. [45] In this study, Caspase-3, Caspase-8, Bcl-2 and Bax gene expression induced by 18-ECFP was quantified in SCC-9 cell line by semi quantitative PCR. On evaluation the expression level of Caspase-3 and Caspase-8 genes by RT-PCR in SCC-9 cells treated with 10 µg/mL and 20 µg/mL of the 18 kDa protein were up regulated when compared to non-treated cells. Bcl-2 was not significantly upregulated. (Table 7–9 and Fig. 12) There was no expression of Bax gene in both control and treated samples. (Fig. 13) The internal control β-actin (housekeeping gene) was used to normalize the Caspase-3, Caspase-8, Bcl-2 and Bax gene expression.

Caspase-3 is a major executioner caspase involved eventually in both intrinsic and extrinsic mechanisms of apoptosis. Caspase – 8 is an initiator caspase which eventually activates executioner Caspase-3 leading to DNA fragmentation. Hence, Caspase-3 and Caspase-8 were chosen for the study.

A recent study by Fiolka et al., 2019, evaluated the apoptotic potential of coelomic fluid of *Dendrobaena veneta* by measuring the level of the activity of re to the test sample by ELISA technique. After incubation at a concentration 250 µg/mL a two-fold increase in the

level of caspase 3, 4, 5 and 10 was witnessed. A small decrease in the activity of initiator caspase 2 was observed. [31] A study by Liu et al., 2017, evaluated the effect of EFE from *Eisenia foetida* on MCF-7 cells. The authors estimated the FAK and CD44v6 expression by RT-PCR and western blotting. They concluded that EFE at 80 µg/mL could inhibit cell adhesion by reducing the protein expression of FAK and CD44v6 in MCF-7 breast cancer cells. [46] Earlier studies have used whole fluids or a single protein component of high concentration of other earthworm specie to evaluate gene expression in cancer cells by PCR. In the current study only 10 and 20 µg/mL concentrations of the protein were used which is an added advantage highlighting the minimal dose required to induce gene expression.

Real time PCR was conducted to validate the results of RT-PCR for Caspase-3 and Caspase-8 gene expression. Q-PCR is a robust and sensitive method to quantitatively assess gene expression in samples. Q-PCR is a reliable molecular technique which has several advantages over conventional RT-PCR.

Q-PCR-based analyses uses the principle of combining the older end-point detection PCR with fluorescent identification methods to report the assimilated amplicons in 'real time' during each PCR cycle when amplified. [47]

Upregulation of Caspase-3 and Caspase-8 in treatment groups in comparison to control was found to be drug concentration dependent. (Table 10 and Fig. 14) Through the Q-PCR results obtained, it is established that the 18-ECFP has upregulated proapoptotic genes Caspase-3 and Caspase-8 in SCC-9 cells.

In the current study, a specific 18 kDa anti-cancer protein of EE was employed and gene regulation was studied on cancer cells *in vitro*. This is the first report in literature of a specific anti-cancer protein from EE upregulating pro-apoptotic genes in oral cancer cells *in vitro* estimated by three different gene expression techniques. This highlights the novelty of the current study. Identification of potent biomolecules through anti-cancer studies may facilitate their usage in drug discovery for adjunctive management of cancer therapies. The mechanism of inhibiting the division and proliferation of cancer cells by these extracts explored could open novel gateways for therapeutic targeting.

5. Conclusion

A novel structure of an unknown anti-cancer protein from ECF has been built for the first-time with molecular dynamics using Schrodinger simulation tools. The *in-silico* analysis of an isolated protein from helminths with cancer receptors has not been described in literature thus far. The results of the current study strongly indicate that the homology model of the 18-ECFP shows satisfactory binding with pro-apoptotic genes Caspase-3 and Caspase-8 responsible for cancer cell death. The above results have been validated through experimental techniques such as RT-PCR and Q-PCR which showed upregulation of Caspase-3 and Caspase-8 genes by the same molecule in oral cancer cell line SCC-9. These findings may lead to the development of naturally available drugs to combat cancer.

The identified biomolecule opens a possibility of its use as a potential cancer chemotherapeutic drug for oral cancer following animal studies and clinical trials in the future. Targeted therapy of oral cancer is promising following identification of anti-cancer biomolecules from ECF with minimal treatment related side effects. Further translational cancer research towards development of anti-cancer drugs to confront oral cancer is the need of the hour. The results presented in this study provide a concrete foundation for further research work in this domain.

Declarations

Acknowledgement

The authors thank Chromous biotech Pvt Ltd for assisting the proteomic procedures. The authors thank Skanda Life Sciences Pvt Ltd (Department for Scientific & Industrial Research recognized laboratory) Bangalore, India for assisting the wet laboratory procedures. The authors thank Pharmacologic Modeling and Simulation Centre (PMSC) of Ramaiah University of Applied Sciences, Bangalore, for assisting the simulation experiments.

Author contributions (names must be given as initials):

(1) Conception and design: DA, RSR, JA, HGNS, SGR, KNCM

(2) Administrative support: DA, RSR, JA

(3) Provision of study materials: DA, HGNS, SGR

(4) Collection and assembly of data: DA, HGNS, SGR

(5) Data analysis and interpretation: DA, JA, HGNS, SGR

(6) Manuscript writing: DA, JA, HGNS, RSR, SGR, KNCM

(7) Final approval of manuscript: All authors

Additional Information (including a Competing Interests Statement): The authors declare that there are no competing interests

References

1. Kalavrezos N, Scully C. (2015) Mouth cancer for clinicians Part 2: epidemiology. *Dental Update*. 42(4):354-9.

2. Vigneswaran N, Williams MD. (2014) Epidemiologic trends in head and neck cancer and aids in diagnosis. *Oral Maxillofac Surg Clin North Am*. 26(2):123-141.
3. Goel A, Singla A, Prabhash K. (2020) Neoadjuvant chemotherapy in oral cancer: Current status and future possibilities. *Cancer Research, Statistics, and Treatment*. 3(1):51.
4. Xu H, Jin X, Yuan Y, Deng P, Jiang L, et al. (2017) Prognostic value from integrative analysis of transcription factors c-Jun and Fra-1 in oral squamous cell carcinoma: a multicentre cohort study. *Scientific reports*. 7(1):1-9.
5. Augustine D, Rao RS, Anbu J, Chidambara Murthy K N. (2017) In vitro antiproliferative effect of earthworm coelomic fluid of Eudrilus eugeniae, Eisenia foetida, and Perionyx excavatus on squamous cell carcinoma-9 cell line: A pilot study. *Phcog Res.* 9, Suppl S1:61-6.
6. Parwanto MLE, Mahyunis, Senjaya H, Edy HJ. (2016) Syamsurizal fractionation and characterization of proteins in Lumbricus rubellus powders. *Int J Pharma Clin Res*, 8(1):15–21.6.
7. Permana S, Hadi RP, Norahmawati E, Endharti AT. (2019) Coelomic fluid of Lumbricus rubellus enhances anti-proliferative effect of 5-fluorouracil by modulating focal adhesion kinase express and IL-1β of colorectal cancer in mice. *J Appl Pharm Sci*. 9(08):041–046.
8. Chen H, Takahashi S, Imamura M, Okutani E, Zhang ZG, et al. (2007) Earthworm fibrinolytic enzyme: Anti-tumor activity on human hepatoma cells in vitro and in vivo. *Chin Med J (Engl)*.120:898-904.
9. Augustine D, Rao RS, Anbu J, Chidambara Murthy KN. (2018) Anticancer prospects of earthworm extracts: A systematic review of in vitro and in vivo studies. *Phcog Rev*, 12:46-55.
10. Augustine D, Rao RS, Jayaraman A, Chidambara Murthy K N. Anti-proliferative activity of earthworm coelomic fluid using oral squamous carcinoma KB 3-1 cells: An In vitro study with serine protease analysis. *Phcog Mag.* 2018; 14:528-34.
11. Augustine D, Rao RS, Anbu J, Chidambara Murthy KN. *In vitro* cytotoxic and apoptotic induction effect of earthworm coelomic fluid of *Eudrilus eugeniae, Eisenia foetida,* and *Perionyx excavatus* on human oral squamous cell carcinoma-9 cell line. *Toxicol Rep*. 2019; 6:347–357.
12. Lin X, Li X, Lin X. A review on applications of computational methods in drug screening and design. *Molecules.* 2020 Jan;25(6):1375.
13. Macalino SJ, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research*. 2015 Sep 1;38(9):1686-701.
14. Augustine D, Rao RS, Anbu J, Saraswathy GR, Chidambara Murthy KN and Hema Sree GNS. Homology Modeling, Molecular Dynamics and Docking of a Novel Anti-cancer Protein Obtained from Earthworm Coelomic Fluid. In: Murahari M, Sundar L, Chaki S, Poongavanam V, Bhat P, Nayak UY, editors. Conference on Drug Design and Discovery Technologies. London, UK, Royal Society of Chemistry; 2019.p.249-253.
15. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019 Jun;18(6):463-477. doi: 10.1038/s41573-019-0024-5.
16. Ambesi-Impiombato A, Bernardo D. Computational biology and drug discovery: From single-target to network drugs. Current Bioinformatics. 2006;1(1):3-13.
17. Sastry, G.M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments," *J. Comput. Aid. Mol. Des,* 2013, 27(3), 221-234.
18. Singhal N, Kumar M, Kanaujia PK, Virdi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol. 2015 Aug 5;6:791. doi: 10.3389/fmicb.2015.00791.
19. Gilany K, Moens L, Dewilde S. Mass spectrometry-based proteomics in the life sciences: a review. *Journal of Paramedical Sciences* (JPS). 2009;1(1):1735-7802.
20. Donnelly DP, Rawlins CM, DeHart CJ, Fornelli L, Schachner LF, Lin Z, et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. Nat Methods. 2019 Jul;16(7):587-594. doi: 10.1038/s41592-019-0457-0.
21. Han X, Aslanian A, Yates JR 3rd. Mass spectrometry for proteomics. Curr Opin Chem Biol. 2008;12(5):483-90. doi: 10.1016/j.cbpa.2008.07.024.
22. Parker CE, Warren MR, Mocanu V. Mass Spectrometry for Proteomics. In: Alzate O, editor. Neuroproteomics. Boca Raton (FL): CRC Press/Taylor & Francis; 2010. Chapter 5. Available from: https://www.ncbi.nlm.nih.gov/books/NBK56011/
23. Banerjee S, Mazumdar S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. Int. J. Anal. Chem. 2012;2012:282574.
24. Pisamai S., Roytrakul S., Phaonakrop N., Jaresitthikunchai J. and Suriyaphol G., (2018). Proteomic analysis of canine oral tumor tissues using MALDI-TOF mass spectrometry and in-gel digestion coupled with mass spectrometry (GeLC MS/MS) approaches. *PLoS ONE*, 13(7), pp.e0200619.
25. Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. Ann Appl Stat. 2010;4(4):1797-1823. doi: 10.1214/10-AOAS341. PMID: 21593992; PMCID: PMC3095207.
26. Rathod RH, Chaudhari SR, Patil AS, Shirkhedkar AA. Ultra-high-performance liquid chromatography-MS/MS (UHPLC-MS/MS) in practice: analysis of drugs and pharmaceutical formulations. *Future Journal of Pharmaceutical Sciences.* 2019;5(1):6.
27. Harvey DJ, Scarff CA, Edgeworth M, Struwe WB, Pagel K, Thalassinos K, Crispin M, Scrivens J. Travelling-wave ion mobility and negative ion fragmentation of high-mannose N-glycans. *Journal of mass spectrometry*. 2016 Mar;51(3):219-35.
28. Chawla G, Ranjan C. Principle, instrumentation, and applications of UPLC: a novel technique of liquid chromatography. *Open Chemistry Journal*. 2016;3(1): 1-16.
29. Wahab MF, O'Haver TC, Gritti F, Hellinghausen G, Armstrong DW. Increasing chromatographic resolution of analytical signals using derivative enhancement approach. *Talanta*.2019;192:492-9.

30. Ciocan-Cartita CA, Jurj A, Buse M, Gulei D, Braicu C, Raduly L et al. The relevance of mass spectrometry analysis for personalized medicine through its successful application in cancer “omics”. *Int. J. Mol. Sci.* 2019; 20(10):2576.
31. Fiołka MJ, Rzymowska J, Bilska S, Lewtak K, Dmoszyńska-Graniczka M, Grzywnowicz K et al. Antitumor activity and apoptotic action of coelomic fluid from the earthworm Dendrobaena veneta against A549 human lung cancer cells. *Apmis.* 2019;127(6):435-48.
32. Abdelmonsef AH, Dulapalli R, Dasari T, Padmarao LS, Mukkera T, Vuruputuri U. Structure based drug discovery of Rab38 protein-identification of antagonists as cancer drug candidates. *Comb Chem High Throughput Screen*. 2016;19:875-92.
33. Salam, NK., Adzhigirey M, Sherman W, Pearlman DA. "Structure-based approach to the prediction of disulfide bonds in proteins," *Protein Eng. Des. Sel.* 2014, 27(10), 365-74.
34. Srinivasan P, Chella Perumal P, Sudha A. Discovery of novel inhibitors for Nek6 protein through homology model assisted structure based virtual screening and molecular docking approaches. *The Scientific World Journal.* 2014 Jan 1;2014.
35. Nagpal N, Goyal S, Dhanjal JK, Ye L, Kaul SC, Wadhwa R, Chaturvedi R, Grover A. Molecular dynamics-based identification of novel natural mortalin–p53 abrogators as anticancer agents. *Journal of Receptors and Signal Transduction.* 2017 Jan 2;37(1):8-16.
36. Zhu K, Day T, Warshaviak D, Murrett C, Friesner R, Pearlman D. Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. *Proteins: Structure, Function, and Bioinformatics*. 2014;82(8):1646-55.
37. Beard H, Cholleti A, Pearlman D, Sherman W, Loving KA. Applying physics-based scoring to calculate free energies of binding for single amino acid mutations in protein-protein complexes. *PloS one*. 2013 Dec 10;8(12):e82849.
38. Sirin S, Pearlman DA, Sherman W. Physics-based enzyme design: Predicting binding affinity and catalytic activity. *Proteins*, 2014, 82(12), 3397-409.
39. Tahlan S, Kumar S, Ramasamy K, Lim SM, Shah SA, Mani V, Narasimhan B. In-silico molecular design of heterocyclic benzimidazole scaffolds as prospective anticancer agents. *BMC chemistry.* 2019;13(1):1-22.
40. Yadav M., Dhagat S. and Eswari J.S. Structure Based Drug Design and Molecular Docking Studies of Anticancer Molecules Paclitaxel, Etoposide and Topotecan using Novel Ligands, *Current Drug Discovery Technologies*. 2020;17(2):183-190
41. Fernández-Recio J, Sternberg MJ. The 4th meeting on the Critical Assessment of Predicted Interaction (CAPRI) held at the Mare Nostrum, Barcelona. *Proteins: Structure, Function, and Bioinformatics*. 2010 Nov 15;78(15):3065-6.
42. Li J, Abel R, Zhu K, Cao Y, Zhao S, Friesner RA. The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics*. 2011 Oct;79(10):2794-812.
43. Lokhande KB, Nagar S, Swamy KV. Molecular interaction studies of Deguelin and its derivatives with Cyclin D1 and Cyclin E in cancer cell signaling pathway: The computational approach. *Scientific reports*. 2019;9(1):1-3.
44. Suganya M, Gnanamangai BM, Ravindran B, Chang SW, Selvaraj A, Govindasamy C, Elsadek MF, Ponmurugan P. Antitumor effect of proanthocyanidin induced apoptosis in human colorectal cancer (HT-29) cells and its molecular docking studies. *BMC chemistry*. 2019; 13(1):1-4.
45. Mullis K, Faloona F, Scharf S, Saiki RK, Horn GT, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In Cold Spring Harbor symposia on quantitative biology 1986 Jan (Vol. 51, pp. 263-273). Cold Spring Harbor Laboratory Press.
46. Liu CM, Chen XT, Pan YY, Liang H, Song SL, Ji AG. Antitumor studies of earthworm fibrinolytic enzyme component a from eisenia foetida on breast cancer cell line MCF-7. *Indian Journal of Pharmaceutical Sciences*. 2017;79(3):361-8.
47. Valasek MA, Repa JJ. The power of real-time PCR. *Adv Physiol Educ*. 2005;29(3):151-159. doi:10.1152/advan.00019.2005

Figures

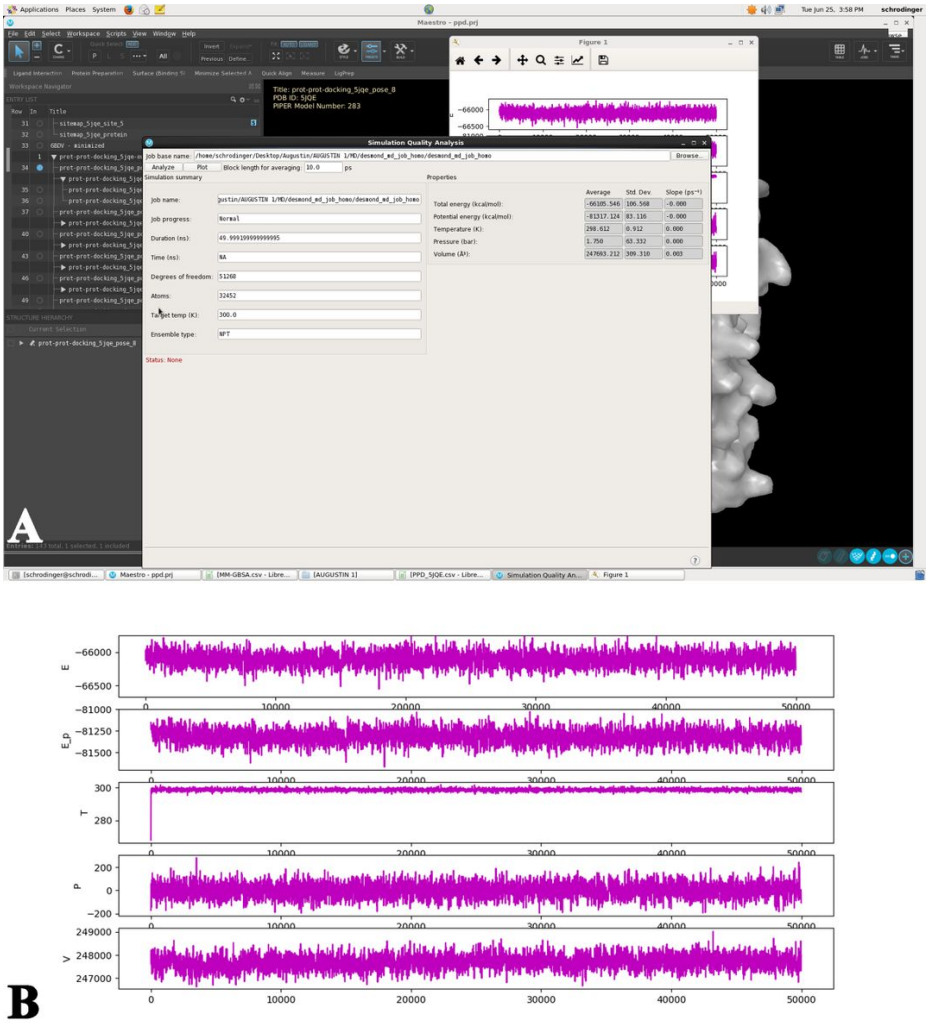


Figure 1

A. Simulation Quality Analysis. B. Simulation Energy Analysis of the Modelled Protein.

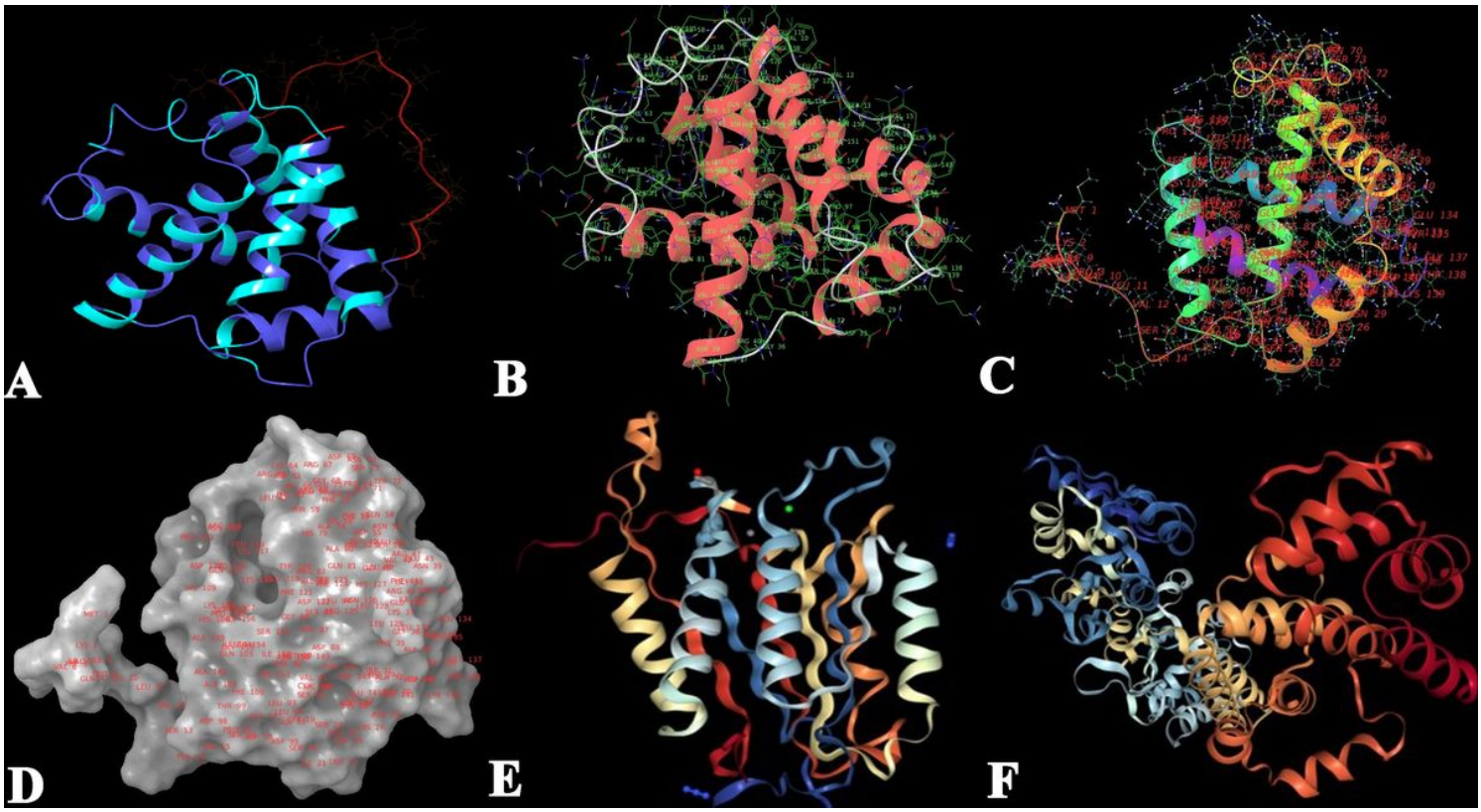


Figure 2

A. Modelled Protein Prior to Docking. B & C. Modelled Protein Prior to Docking – with Amino Acid Details. D. Fluid Model View of Modelled Protein Prior to Docking. E. Structure of Caspase - 3 Receptor (PDB:6BDV). F. Structure of Caspase - 8 Receptor (PDB: 5JQE).

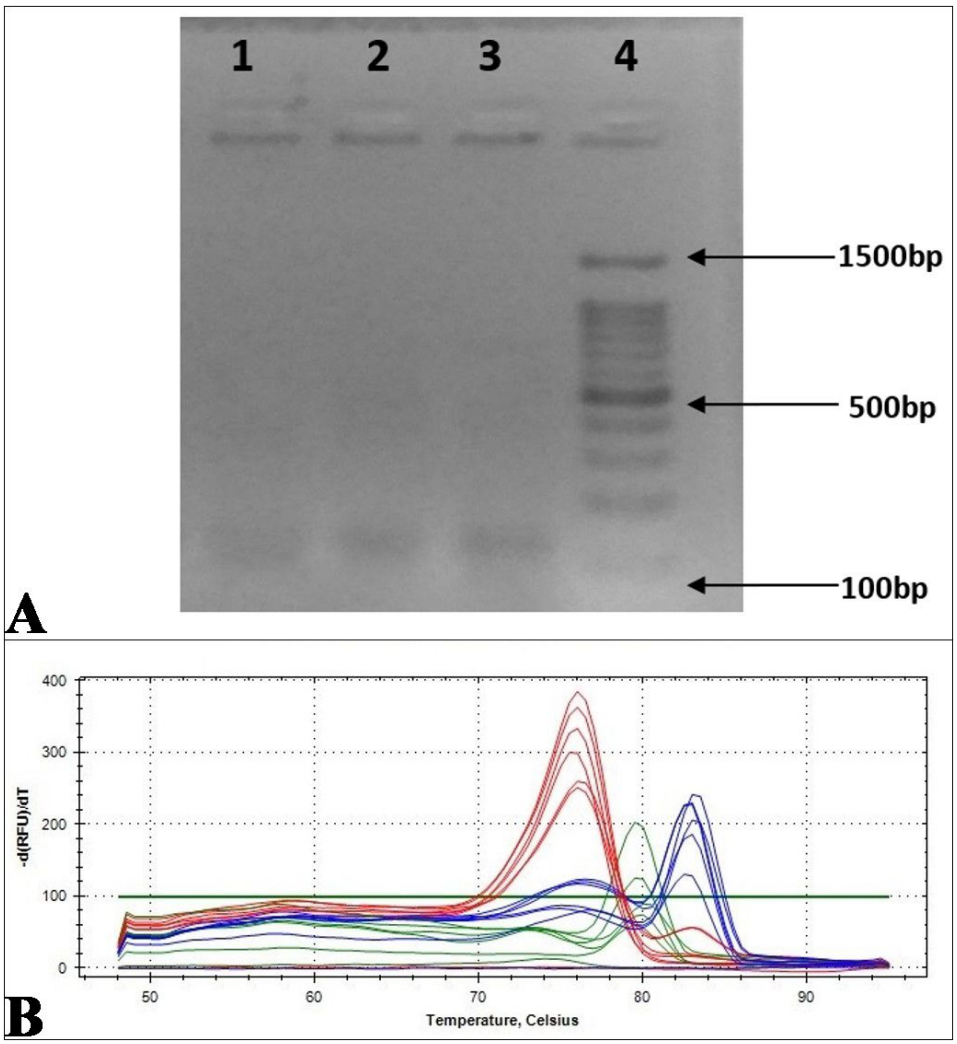
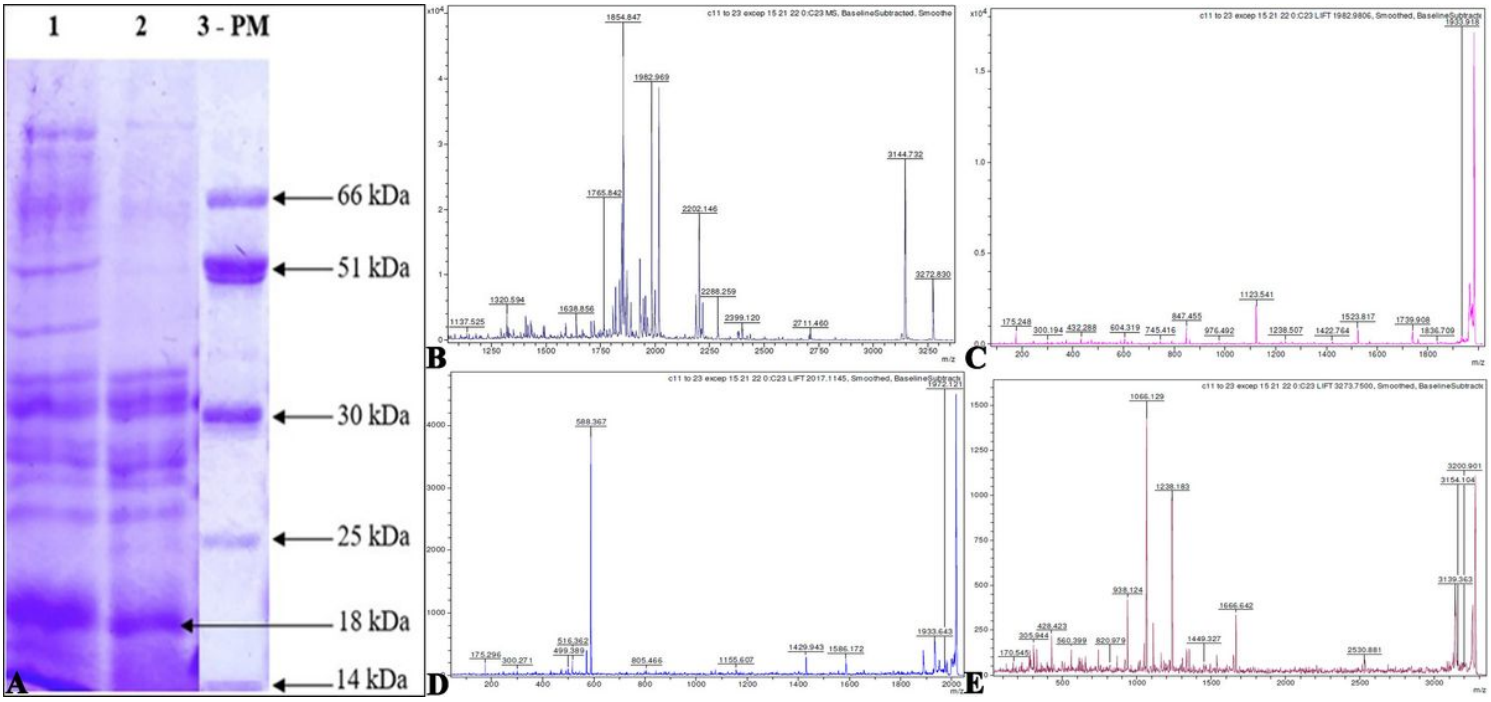


Figure 3

A. RT-PCR, Amplification of β-Actin Gene in SCC-9 Cells. Lane 1-Sample Treatment 20 µg/ml; Lane 2- Sample Treatment 10µg /ml; Lane 3- Control Untreated Cells; Lane 4-100 BP Marker. B. Q-PCR, Melt Peak Profiling of Caspase-3, Caspase-8 and GAPDH Genes.



A. SDS PAGE Analysis Depicting the Separated 18 kDa Protein in Lane 2. B. PMF of the Trypsin Digested Mixture by MALDI-TOF-Mass Spectrometric Analysis. C, D & E. Peptide Analysis Graphs Obtained through MS/MS Sequencing.

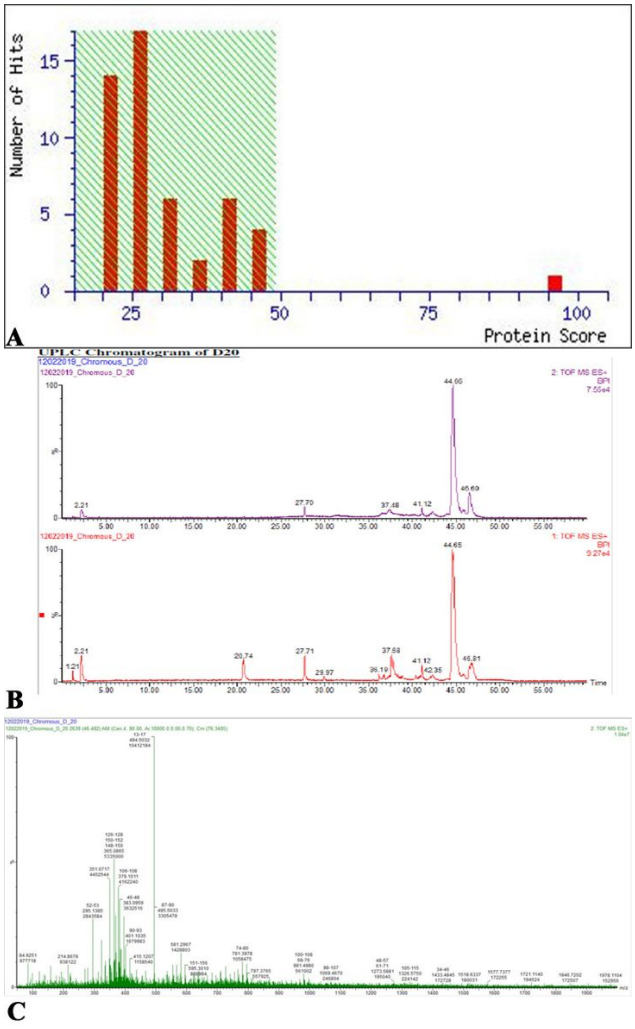


Figure 5

A. Mascot Score Histogram, Protein Report B. The UPLC Chromatogram of the 18-ECFP. C. MS/MS spectrum of 18-ECFP.

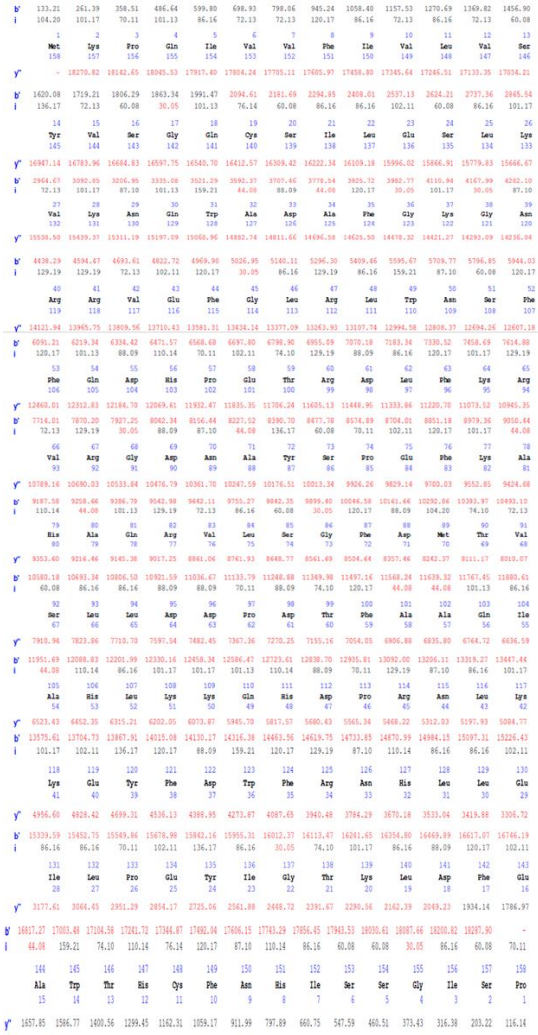


Figure 6

AA-Sequence of 18-ECFP.

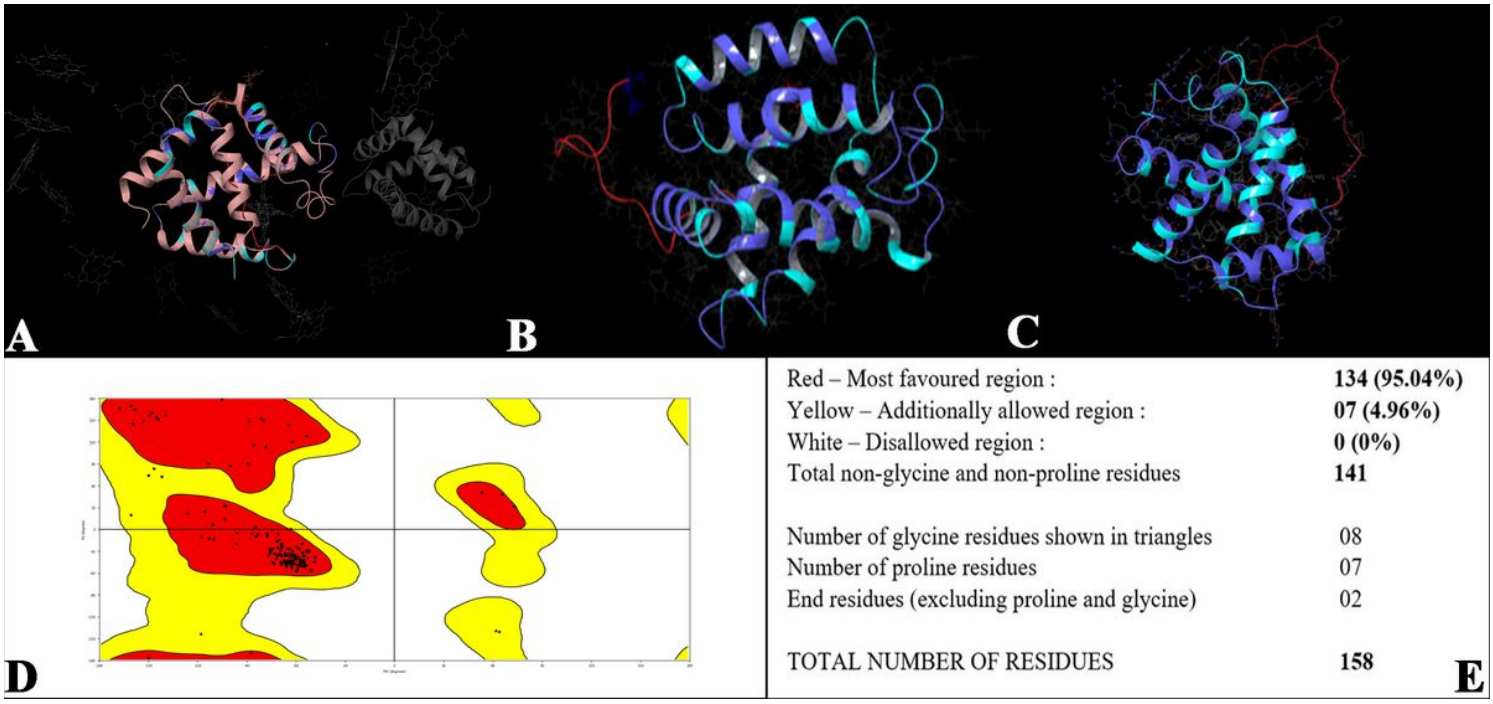


Figure 7

A. The Unrefined Homology Model of 18-ECFP. B & C. Refined Homology Model of 18-ECFP after Protein Preparation. (Colours: Cyan – Positives, Blue – Identifiers, Red – Insertions. Shapes: Ribbon – β – Sheet, Arrow – α – Helix, Loop – B-Loop/Bend/Turn). D & E. Ramachandran Plot for Validation of Homology Model.

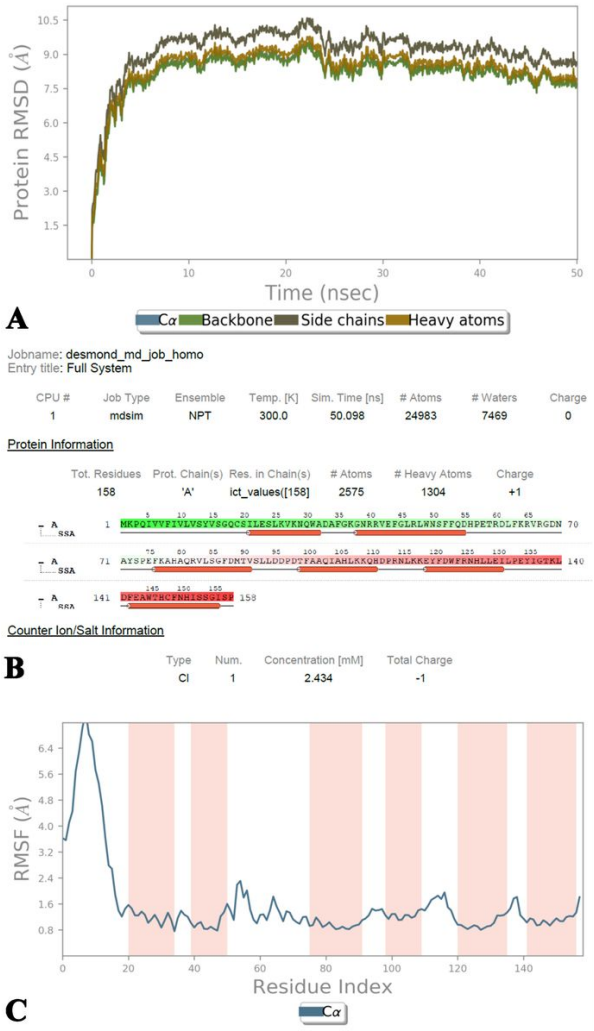


Figure 8

A. Protein-Ligand RMSD. B. Simulation Interactions Diagram Report C. Protein RMSF of the 18 kDa protein.

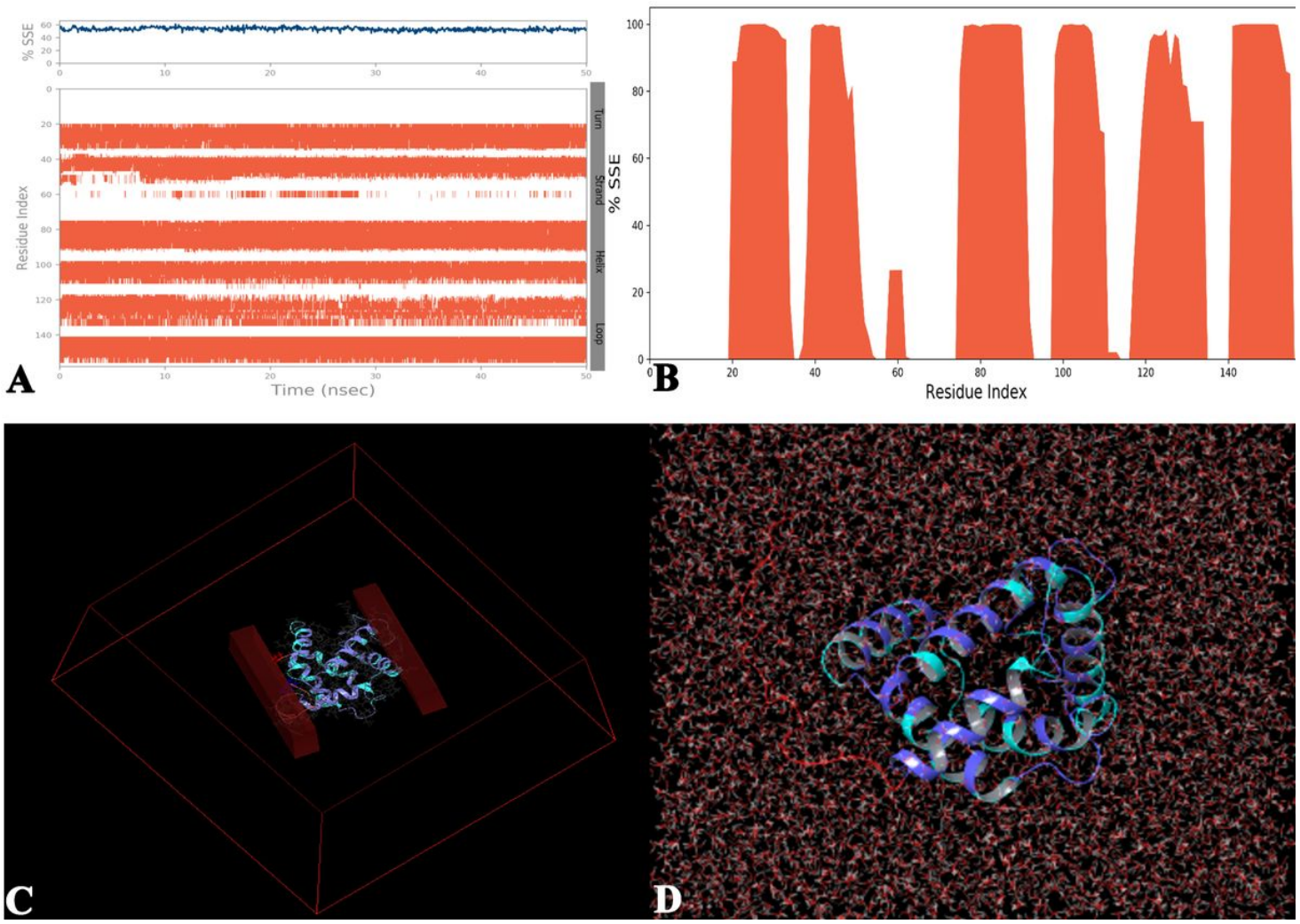


Figure 9

A. The SSE Graph with Alpha-Helical and Beta-Strand Regions Highlighted. B. SSE Distribution by Residue Index. C & D. Modelled Protein in its orthogonal matrix.

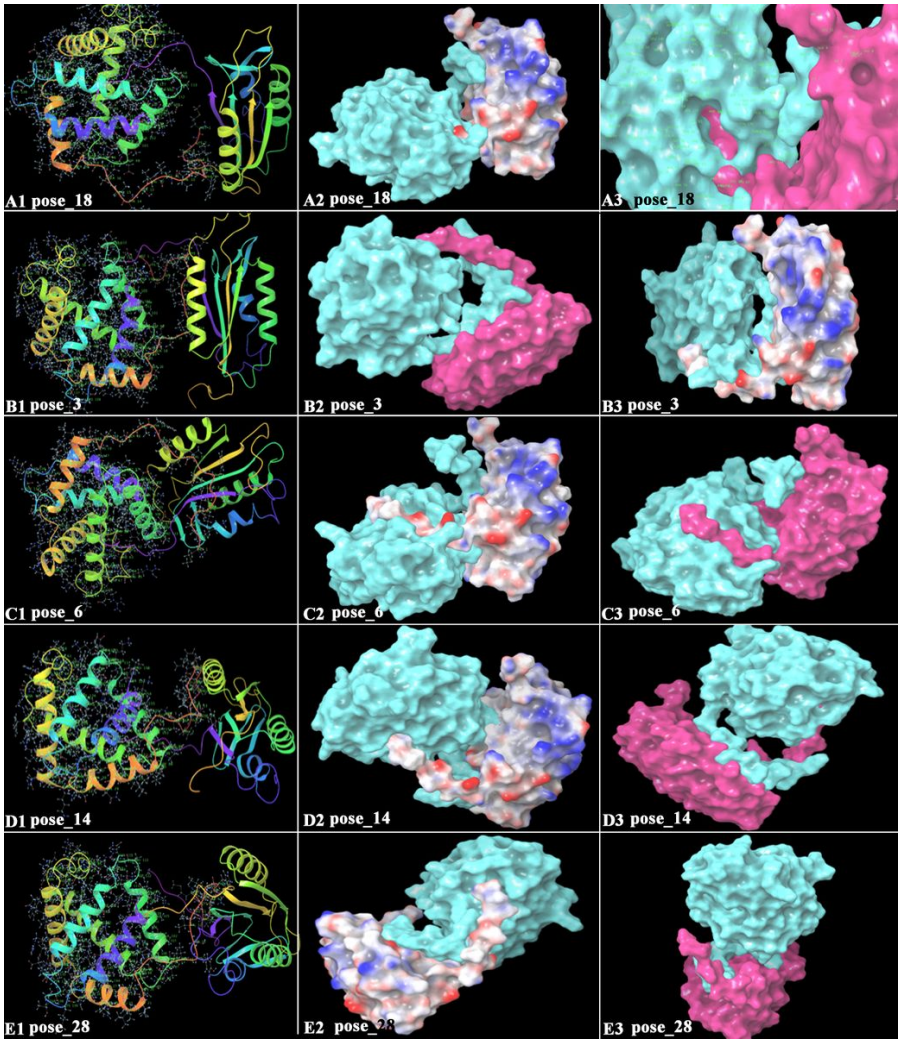


Figure 10

PPD between 18-ECFP (cyan) with Caspase-3 Receptor (PDB:6BDV). 5 Poses with Highest PIPER SCORE, A1-3. Pose_18. B1-3. Pose_3. C1-3. Pose_6. D1-3. Pose_14. E1-3. Pose_28.

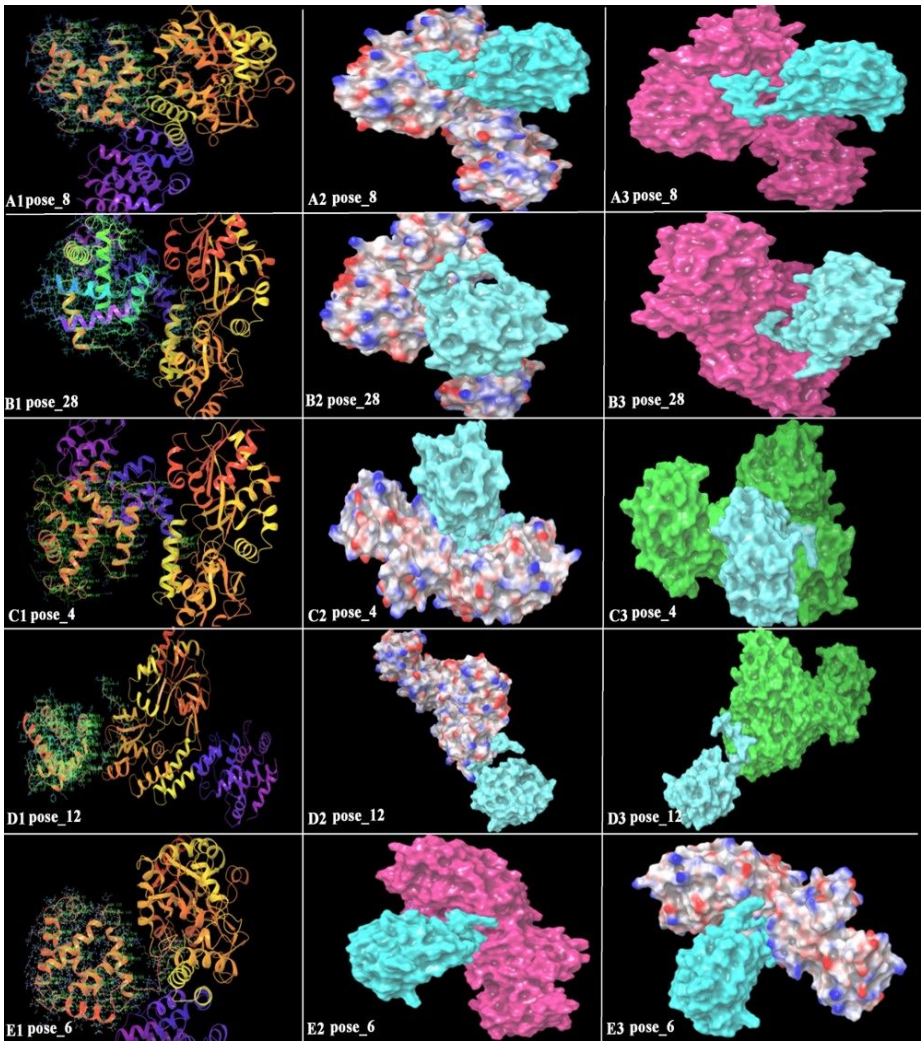


Figure 11

PPD between 18-ECFP (cyan) with Caspase-3 Receptor (PDB:6BDV). 5 Poses with Highest PIPER SCORE, A1-3. Pose_18. B1-3. Pose_3. C1-3. Pose_6. D1-3. Pose_14. E1-3. Pose_28.

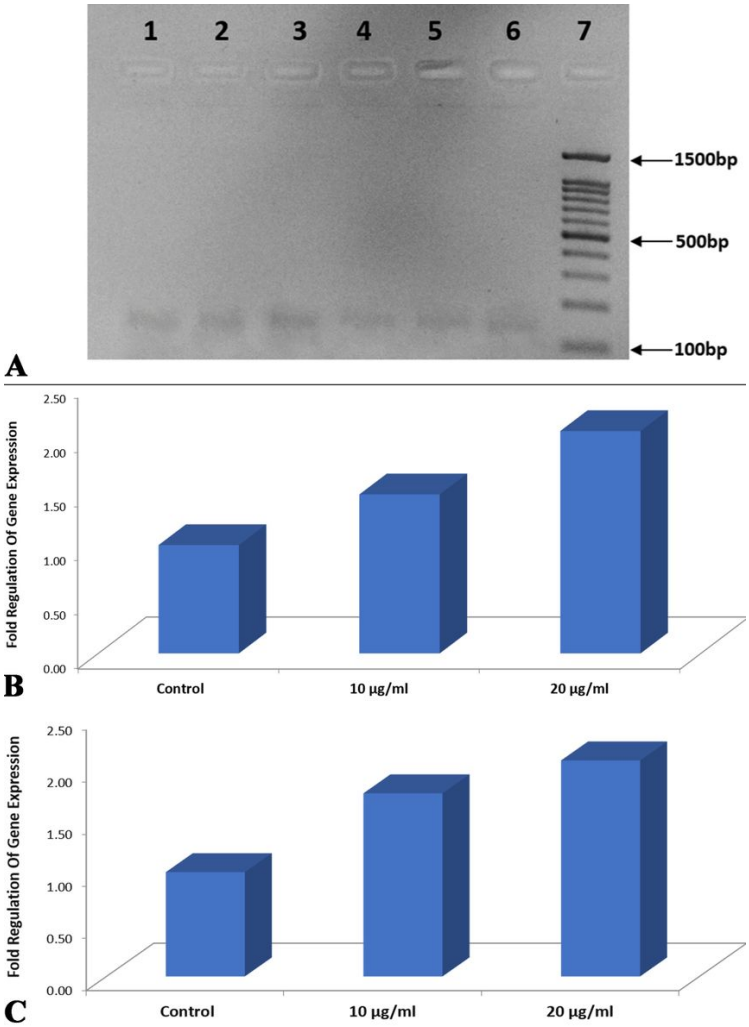


Figure 12

A. Amplification of Caspase-8 and Caspase-3 Gene in SCC-9 Cells. Lane 1- Control Untreated Cells (Caspase-3); Lane 2-Sample Treatment 10 µg/ml (Caspase-3); Lane 3- Sample Treatment 20 µg/ml (Caspase-3); Lane 4- Control Untreated Cells (Caspase-8); Lane 5-Sample Treatment 10 µg/ml (Caspase-8); Lane 6-Sample Treatment 20 µg/ml (Caspase-8); Lane 7-100 BP Marker. B. Relative Expression of Caspase-3 Gene in Untreated and Treated SCC-9 Cells. C. Relative Expression of Caspase-8 Gene in Untreated and Treated SCC-9 Cells.

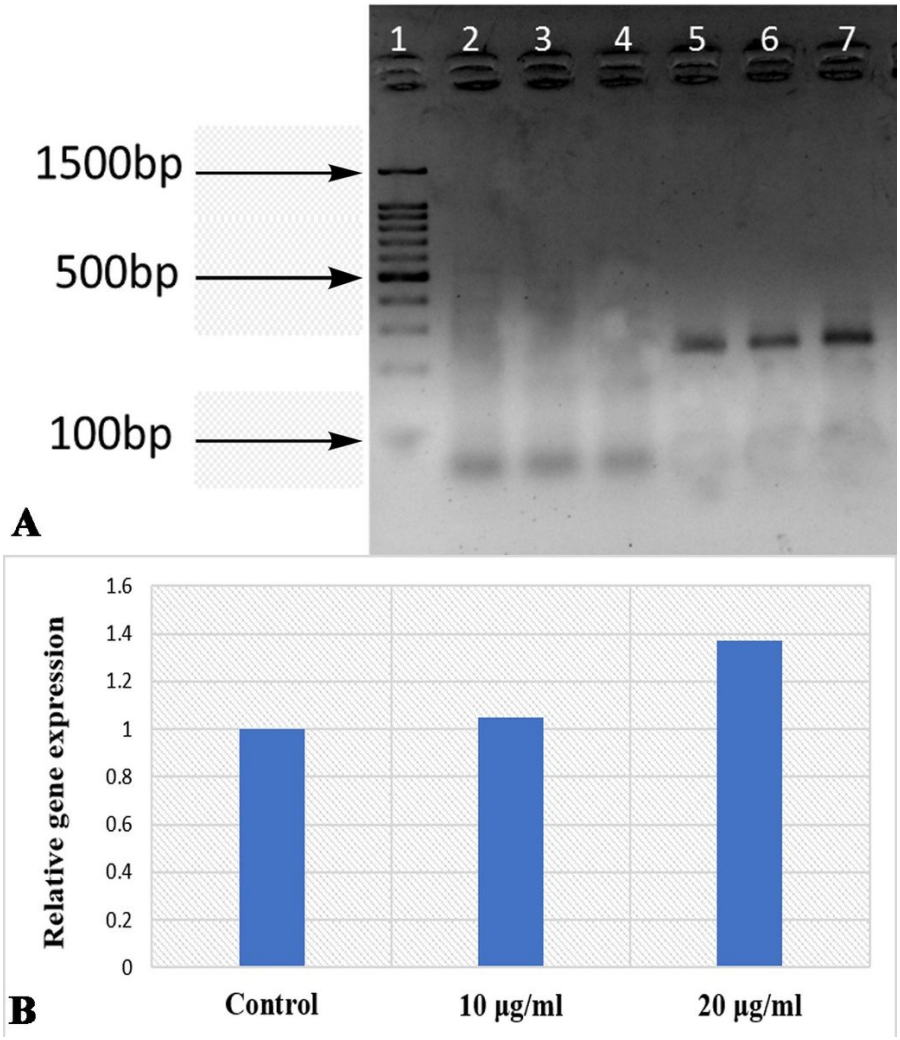


Figure 13

A. Amplification of Bcl-2 Gene in SCC-9 Cells. Lane 1-100 BP Marker; Lane 2- Control Untreated Cells (Bax); Lane 3-Sample Treatment 10 µg/ml (Bax); Lane 4-Sample Treatment 20 µg/ml (Bax); Lane 5- Control Untreated Cells (Bcl-2); Lane 6-Sample Treatment 10 µg/ml (Bcl-2); Lane 7-Sample Treatment 20 µg/ml (Bcl-2). Bax Gene was not expressed. B. Relative Expression of Bcl-2 Gene in Untreated and Treated SCC-9 Cells.

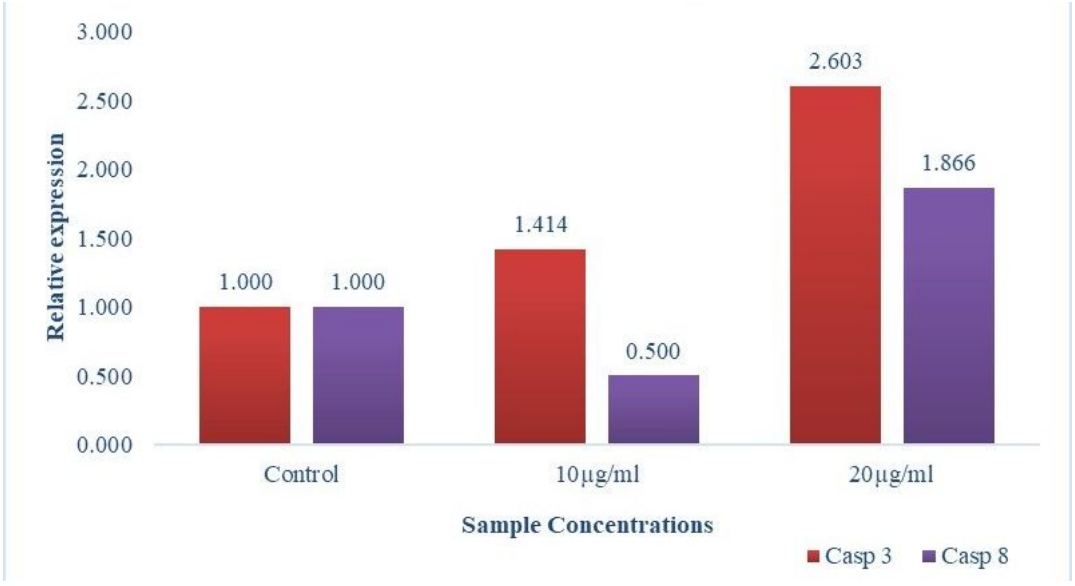


Figure 14

Q-PCR Relative Expression of Caspase-3 and Caspase-8 in Untreated and Treated SCC-9 Cells.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Flowchart1.docx
- Appendix1.docx