Section A-Research paper

MOLECULAR DOCKING OF BIOACTIVE COMPOUNDS AGAINST p38, EGFR AND BCL-2 PROTEINS

[1] Abdulathim Alsadiq Dhaw Tkiwiriyah, [2*] Asita Elengoe, [3] Faraj Khamees Saqar

[1,2] Department of Biotechnology, Faculty of Applied Science, Lincoln University College, 47301 Petaling Jaya, Selangor, Malaysia

[1,3] Department of Medical Laboratories, Faculty of Medical Technology, Azzaytuna University, 00218 Tarhuna , Libya

*Corresponding author email: asitaelengoe@yahoo.com

ABSTRACT

In this study, the three-dimensional (3-D) structures of lung cancer cell line proteins (p38, epidermal growth factor (EGFR) and B-cell lymphoma 2 (Bcl-2), were developed and their binding affinities with coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin through local docking were evaluated. Firstly, Swiss model was used to generate p38, EGFR and Bcl-2. After that, they were viewed by the PyMol software. Next, Expasy's ProtParam Proteomics server was used to determine the physical and chemical parameters of the protein models. Moreover, the protein models validated using PROCHECK, ProQ, ERRAT and Verify 3D programs. Finally, the protein models were docked with coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin by using BSP-Slim server. After the validation and verification process, all the protein models were stable. The p38 showed binding energy with coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin at -6.72, -5.97, -6.98, -6.52, -7.70 and -7.21 kcal/mol respectively. Besides that, the EGFR showed binding energy of coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin at -7.38, -6.53, -7.41, -7.05, -6.69 and -6.84 kcal/mol respectively. Moreover, the Bcl-2 showed binding energy of coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin at -6.74, -5.97, -6.06, -5.41, -6.54 and -5.13 kcal/mol respectively. The p38 had the strongest bond with wogonin; the EGFR protein had the strongest interaction with gingerol; while Bcl-2 had the strongest binding affinity with coumaric acid.

Keywords: p38, EGFR, Bcl-2, Coumaric Acid, Anthraquinone, Gingerol, Tanshinone, Baicalein, Wogonin, Docking

Introduction

Cancer is the second-leading cause of death worldwide, with an anticipated 9.6 million deaths in 2018. Around one-third of cancer deaths are due to the five main behavioral and dietary risks: high body mass index, poor consumption of fruits and vegetables, lack of physical

activity, and use of tobacco and alcohol. Tobacco use is the most significant risk factor for cancer and is responsible for about 22% of deaths from cancer. Global findings indicate that almost half of the new cases and more than half of the world's cancer deaths in 2018 are expected to occur in Asia for men and women combined, partially because the country has nearly 60 percent of the global population. The most common cancers are lung (2.09 million cases), breast (2.09 million cases), prostate (1.28 million cases), skin cancer (non-melanoma) (1.04 million cases) and stomach (1.03 million cases). Some reports on lung cancer include both small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). In total, SCLC is around 13% of all lung cancers, and NSCLC is about 84% (Sakiyama et al. 2005).

The proteins that can be found in lung cancers are p38, EGFR, and Bcl-2 proteins. p38, a MAPK normally associated with stress response, growth arrest, and apoptosis, was triggered in all human lung cancer samples, indicating an additional role in malignant cell growth or transformation for this pathway. The pathways of mitogen-activated protein kinase (MAPK) relay signals from the cell membrane to the nucleus. MAPK cascade activation may play a part in malignant transformation. Beckles and his co-researchers hypothesized that there will be increased expression of one or more of those pathways in human lung cancers (Beckles et al. 2003). Using Western blot analysis of the homogeneous tissue from resected non-small cell lung cancers and matched non-neoplastic lung tissue, they found that only activated p38 was consistently increased in tumors relative to normal tissue. *In vitro,* kinase assays indicated that the activated MAPK levels were associated with the enzyme activity, and the immunohistochemical analysis indicated the activated MAPKs' cell localization. In a hypoxic chamber, we incubated a line of lung cancer cells to replicate the hypoxic condition in solid lung tumors but found no increase in activation of p38 (Crane et al. 2016).

Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein with a binding domain of the extracellular epidermal growth factor and an intracellular tyrosine kinase domain that regulates signaling pathways to control cell proliferation. The EGFR results in autophosphorylation by intrinsic tyrosine and activating multiple cascades of signal transduction through binding to its ligand (Mao et al. 1994). Constitutive or prolonged activation of these downstream target sequences is thought to create more aggressive phenotypes of tumors. Mutations in EGFR have been discovered in association with some lung cancers. In a retrospective study, the expression of the proto-oncogene Bcl-2, whose principal role tends to be apoptosis inhibition, was examined in 164 cases of primary small-cell lung cancer by immunohistochemistry. Bcl-2 expression was demonstrated in one hundred twenty-five cases (76%). There was no difference between the two groups regarding serum lactate dehydrogenase (LDH) levels and proliferative activity. A study found that patients with Bcl-2 positive tumors had a median survival time of 12 months compared to 9.5 months for patients with Bcl-2 negative tumors (Crane et al. 2015).

There is an urgent need to find better care for this disease. Since chemotherapy and radiation therapy cause numerous side effects, it is therefore important to discover new agents for the treatment of this disease (Blair et al. 1983). Plants have also been the foundation of conventional medical systems. They have been supplying mankind with continuous remedies

for thousands of years. Knowledge of medicinal plants for the preparation of various medicines was of great importance (Rami-Porta et al. 2007). Medicinal plants are known as a rich source of a wide variety of ingredients (plant compounds) that can be used for drug production (Tockman 2008). The biological activities of coumaric acid, including antioxidant, anti-cancer, anti-microbial, anti-virus, anti-inflammatory, anti-platelet aggregation, anxiolytic, antipyretic, analgesic, and anti-arthritis, anti-diabetes, anti-obesity, anti-hyperlipaemia and anti-gout. The naturally occurring anthraquinones possess a broad spectrum of bioactivities, such as cathartic, anti-cancer, anti-inflammatory, anti-microbial, diuretic, vasorelaxant, and phytoestrogen activities, suggesting their possible clinical application in many diseases. Gingerol finds in the rhizomes of ginger (*Zingiber officinale* Roscoe). It possesses medicinal properties such as relieving pain; alleviating nausea, anti-arthritis, anti-oxidant, anti-inflammatory, etc. Tanshinone is a natural compound extracted from *Salvia miltiorrhiza*. Tanshinones offer a variety of pharmacological properties, including antioxidant, anti-inflammatory, anti-cancer, phytoestrogenic action, vasodilation, neuroprotection, and regulation of metabolic function. A flavone, or type of flavonoid, called baicalein (5,6,7-trihydroxyflavone), was first discovered in the roots of *Scutellaria baicalensis* and *Scutellaria lateriflora*. It can trigger cell cycle arrest, induces apoptosis, and prevents signal pathways. It has anti-cancer, anti-oxidant, and anti-inflammatory activities. In *Scutellaria baicalensis*, a flavonoid-like chemical molecule known as wogonin is an O-methylated flavone. Wogonin, which has been used to treat lung inflammation, offers potential benefits in inflammatory colorectal carcinogenesis, particularly through the reduction of NF-B and activation of Nrf2 signaling pathways in HCT116 cells and THP-1 cells (Collins et al. 2007).

The study aims to generate three-dimensional (3-D) models of the target proteins (lung cancer cell proteins such as p38, EGFR and Bcl-2) and then determine their binding affinities with coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin respectively via local docking approach.

Methodology

Target Sequence

The complete amino acid sequence of the p38 (GI: 13124003), EGFR (GI: 110002567) and Bcl-2 (GI: 72198346) were obtained from National Center for Biotechnology Information (NCBI). p38, EGFR and Bcl-2 contain 338, 464 and 205 amino acids respectively.

Homology modeling

The 3-D models of p38, EGFR and Bcl-2 were generated using the Swiss Model based on their relative amino acid sequences (Biasini et al. 2014) and visualized by the PyMol software (DeLano 2004; Rigsby and Parker 2016).

Physiochemical characterization

Physiochemical characterization analysis was carried out via the Expasy's ProtParam Proteomics server (Kalet 2014). The ESBRI program was used to discover salt bridges in the protein models (Verma and Singh 2013); the number of disulfide bonds was calculated using the Cys_Rec program (Costantini et al. 2008).

Prediction of the secondary structures

The secondary structural features were predicted using Self-Optimized Prediction Method from Alignment (SOPMA) (Angamuthu and Piramanayagam, 2017).

Evaluation of the protein models

The protein structures were validated with PROCHECK by Ramachandran plot analysis (Tiwari et al. 2016). The model was further analyzed by ProQ (Uziela et al. 2016), ERRAT (Appaiah and Vasu 2016) and Verify 3D programs ((Monie et al. 2016).

Identification of active sites

To identify the binding sites of the p38, EGFR, and Bcl-2, they were submitted to an active site-prediction server (SCFBio) (Singh et al. 2011).

Ligand preparation

The complete sequences of the six ligand models (coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin) were obtained from the PubChem database and converted into Protein Data Bank (PDB) format as the structures were in structured data format (Kasilingam and Elengoe 2018).

Docking of the target protein and plant compound

The 3-D models of p38, EGFR and Bcl-2 proteins were performed for docking with the plant compounds (coumaric acid, anthraquinone, gingerol, tanshinone, baicalein and wogonin) using a BSP-Slim server. The best score from the docking was selected (Supramaniam and Elengoe 2020; Elengoe and Sebestian 2020).

Result and Discussion

Physiochemical characterization

The total number of amino acids for p38, EGFR, and Bcl-2 are 338, 464 and 205 amino acids respectively. Moreover, the molecular weight of p38, EGFR, and Bcl-2 proteins are 38274.32, 50343.70 and 22337.18 Daltons respectively. On the other hand, the isoelectric point (pI) of p38 and Bcl-2 are less than 7.10, indicating acidic characteristics. However, the isoelectric point (pI) of EGFR is more than 7, indicating alkaline characteristics. Generally, the isoelectric point (pI) of a protein is the key characteristic that influences its overall electrostatics behavior (Lapinska et al. 2017). Next, the amount of negatively charged residues (Asp+Glu) and positively charged residues (Arg+Lys) for p38 are 51 and 42

respectively. While EGFR has 49 negatively charged residues and 49 positively charged residues whereas Bcl-2 has 18 negatively charged residues and 20 positively charged residues. Furthermore, the instability index of p38, EGFR and Bcl-2 are computed to be 23.51, 35.56, and 53.72. Thus, p38 and EGFR are classified as stable while Bcl-2 is unstable. Table 1 shows the physiochemical characterization of EGFR, p38 and Bcl-2 proteins obtained from ExPASy ProtParam tool.

Table 1: Physiochemical characterization of EGFR, p38 and Bcl-2 proteins obtained from ExPASy ProtParam tool

Protein / **Parameters**	**p38**	**EGFR**	**Bcl-2**
GI	13124003	110002567	72198346
Number of amino acids	338	464	205
Molecular weight	38274.32	50343.70	22337.18
Theoretical pI	5.41	7.10	6.43
Amino acid composition	Ala (A) 20 5.9% Arg (R) 15 4.4% Asn (N) 14 4.1% Asp (D) 14 4.1% Cys (C) 3 0.9% Gln (Q) 11 3.3% Glu (E) 37 10.9% Gly (G) 26 7.7% His (H) 6 1.8% Ile (I) 15 4.4% Leu (L) 24 7.1% Lys (K) 27 8.0% Met (M) 8 2.4% Phe (F) 15 4.4% Pro (P) 16 4.7% Ser (S) 17 5.0% Thr (T) 29 8.6% Trp (W) 10 3.0% Tyr (Y) 7 2.1% Val (V) 24 7.1% Pyl (O) 0 0.0% Sec (U) 0 0.0% (B) 0 0.0% (Z) 0 0.0% (X) 0 0.0%	Ala (A) 25 5.4% Arg (R) 22 4.7% Asn (N) 26 5.6% Asp (D) 22 4.7% Cys (C) 42 9.1% Gln (Q) 13 2.8% Glu (E) 27 5.8% Gly (G) 45 9.7% His (H) 15 3.2% Ile (I) 27 5.8% Leu (L) 34 7.3% Lys (K) 27 5.8% Met (M) 7 1.5% Phe (F) 11 2.4% Pro (P) 28 6.0% Ser (S) 26 5.6% Thr (T) 28 6.0% Trp (W) 4 0.9% Tyr (Y) 9 1.9% Val (V) 26 5.6% Pyl (O) 0 0.0% Sec (U) 0 0.0% (B) 0 0.0% (Z) 0 0.0% (X) 0 0.0%	Ala (A) 24 11.7% Arg (R) 16 7.8% Asn (N) 6 2.9% Asp (D) 11 5.4% Cys (C) 1 0.5% Gln (Q) 6 2.9% Glu (E) 9 4.4% Gly (G) 21 10.2% His (H) 8 3.9% Ile (I) 6 2.9% Leu (L) 14 6.8% Lys (K) 2 1.0% Met (M) 6 2.9% Phe (F) 9 4.4% Pro (P) 17 8.3% Ser (S) 12 5.9% Thr (T) 10 4.9% Trp (W) 5 2.4% Tyr (Y) 6 2.9% Val (V) 16 7.8% Pyl (O) 0 0.0% Sec (U) 0 0.0% (B) 0 0.0% (Z) 0 0.0% (X) 0 0.0%
Total number of negatively charged residues (Asp+Glu)	51	49	20
Total number of positively charged residues (Arg+Lys)	42	49	18

Atomic composition	Carbon C 1710 Hydrogen H 2657 Nitrogen N 457 Oxygen O 519 Sulfur S 11	Carbon C 2152 Hydrogen H 3434 Nitrogen N 630 Oxygen O 665 Sulfur S 49	Carbon C 997 Hydrogen H 1516 Nitrogen N 288 Oxygen O 286 Sulfur S 7
Formula	$C_{1710}H_{2657}N_{457}O_{519}S_{11}$	$C_{2152}H_{3434}N_{630}O_{665}S_{49}$	$C_{997}H_{1516}N_{288}O_{286}S_{7}$
Total number of atoms	5354	6930	3094
Extinction coefficients	Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water. Ext. coefficient 65555 Abs 0.1% (=1 g/l) 1.713, assuming all pairs of Cys residues form cysteines Ext. coefficient 65430 Abs 0.1% (=1 g/l) 1.710, assuming all Cys residues are reduced	Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water. Ext. coefficient 38035 Abs 0.1% (=1 g/l) 0.756, assuming all pairs of Cys residues form cysteines Ext. coefficient 35410 Abs 0.1% (=1 g/l) 0.703, assuming all Cys residues are reduced	Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water. Ext. coefficient 36440 Abs 0.1% (=1 g/l) 1.631, assuming all pairs of Cys residues form cysteines Ext. coefficient 36440 Abs 0.1% (=1 g/l) 1.631, assuming all Cys residues are reduced
Estimated half-life	The N-terminal of the sequence considered is M (Met). The estimated half-life is: 30 hours (mammalian reticulocytes, *in vitro*). >20 hours (yeast, *in vivo*). >10 hours (*Escherichia coli*, *in vivo*).	The N-terminal of the sequence considered is M (Met). The estimated half-life is: 30 hours (mammalian reticulocytes, *in vitro*). >20 hours (yeast, *in vivo*). >10 hours (*Escherichia coli*, *in vivo*).	The N-terminal of the sequence considered is M (Met). The estimated half-life is: 30 hours (mammalian reticulocytes, *in vitro*). >20 hours (yeast, *in vivo*). >10 hours (*Escherichia coli*, *in vivo*).
Instability index	The instability index (II) is computed to be 23.51 This classifies the protein as stable.	The instability index (II) is computed to be 35.56 This classifies the protein as stable.	The instability index (II) is computed to be 53.72 This classifies the protein as unstable.
Aliphatic index	71.51	72.91	72.39
Grand average of hydropathicity (GRAVY)	-0.550	-0.269	-0.255

Next, Cys_Rec was used to evaluate the position of a cysteine residue, the total number of cysteines present and the pattern of pairs present in the protein sequences as output. Table 2 shows the Cys_Rec result of p38, EGFR and Bcl-2 proteins

Table 2: Cys_Rec result of p38, EGFR and Bcl-2 proteins

Protein	Cys_Rec	Score
p38	CYS_56	-29.8
	CYS_207	-32.3
	CYS_301	-17.9
	CYS_9	-13.0
	CYS_13	39.2
	CYS_17	100.1
	CYS_25	98.3
	CYS_26	104.2
	CYS_30	104.1

EGFR	CYS_34	90.5
	CYS_42	48.2
	CYS_45	56.0
	CYS_54	58.2
	CYS_58	49.5
	CYS_85	55.7
	CYS_89	54.9
	CYS_101	50.4
	CYS_105	44.0
	CYS_120	63.0
	CYS_123	73.3
	CYS_127	75.3
	CYS_131	61.6
	CYS_156	33.0
	CYS_264	45.3
	CYS_293	43.8
	CYS_300	56.5
	CYS_304	46.8
	CYS_309	66.0
	CYS_317	65.6
	CYS_320	60.2
	CYS_329	49.1
	CYS_333	42.2
	CYS_349	42.2
	CYS_352	32.9
	CYS_356	62.9
	CYS_373	54.2
	CYS_376	54.8
	CYS_385	35.8
	CYS_389	41.2
	CYS_411	78.8
	CYS_414	85.1
	CYS_418	84.4
	CYS_422	26.5
	CYS_435	22.5
Bcl-2	CYS_158	-28.9

Apart from that, salt bridges are the specific electrostatic interactions that contribute to the overall stability of proteins in which these interactions play a crucial role in the nucleation process of the hierarchical protein folding model. In this study, the number of salt bridges of p38, EGFR, and Bcl-2 obtained from the ESBRI is 12, 35 and 9 respectively. Table 3 shows the ESBRI result of p38, EGFR and Bcl-2 proteins.

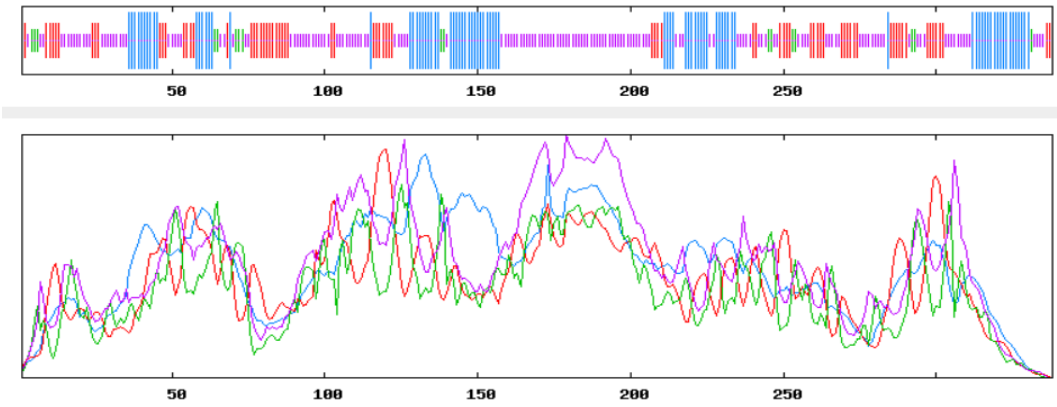
Table 3: ESBRI result of p38, EGFR and Bcl-2 proteins

Proteins	Residue 1	Residue 2	Distance
p38	NZ GLU A 10	OE1 LYS A 125	2.73
	NH1 GLU A 17	OE2 ARG A 21	6.72
	NH1 GLU A 18	OE1 ARG A 21	3.73
	NH1 ASP A 42	OD1 ARG A 43	6.61
	NZ GLU A 57	OE1 LYS A 46	3.79
	NH1 ASP A 60	OD1 ARG A 43	6.40
	NH1 GLU A 78	OE1 ARG A 72	5.21
	NZ GLU A 96	OE1 LYS A 9	6.97
	NZ GLU A 100	OE1 LYS A 5	4.44
	NH1 GLU A 108	OE1 ARG A 111	5.04

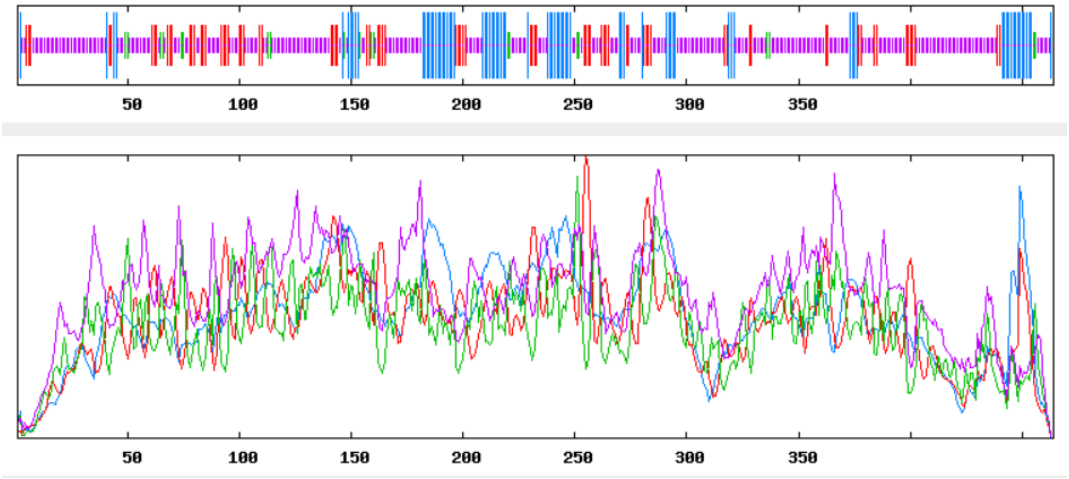
	NH1 GLU A 122	OE1 ARG A 118	5.65
	NZ ALA A 132	O LYS A 125	5.86
EGFR	NH1 ASP A 18	OD2 ARG A 12	6.23
	ND1 GLU A 33	OE1 HIS A 21	5.77
	NH2 ASP A 35	OD2 ARG A 32	6.96
	NZ ASP A 44	OD1 LYS 49	4.64
	NH1 GLU A 45	OE1 ARG A 43	5.32
	NH1 ASP A 50	OD1 ARG A 40	5.14
	ND1 ASP A 91	OD1 HIS A 92	6.00
	NH1 GLU A 105	OE1 ARG A 112	3.07
	NZ ASP A 109	OD1 ARG A 112	6.89
	ND1 GLU A 132	OE1 HIS A 146	5.82
	ND1 ASP A 156	OD1 LYS A 132	6.68
	NZ GLU A 179	OE1 LYS A 145	5.82
	NZ ASP A 181	OD1 LYS 184	4.70
	NH1 GLU A 188	OE1 ARG A 122	4.41
	NZ ASP A 200	OE2 LYS A 206	6.88
	NH1 ASP A 204	OD1 HIS 202	6.65
	NZ GLU A 209	OE1 LYS A 184	6.00
	NZ GLU A 212	OE1 LYS A 187	4.52
	NH1 GLU A 243	OE1 ARG A 215	5.44
	NZ ASP A 246	OD2 LYS 219	6.89
	NZ ASP A 248	OD1 LYS A 288	4.68
	NZ GLU A 284	OE1 LYS A 267	2.50
	NZ GLU A 301	OE1 LYS A 267	2.84
	NH2 GLU A 307	OE1ARG A 239	6.52
	NH1 ASP A 110	OD1 ARG A 239	4.49
	NH1 GLU A 322	OE1 ARG A 309	3.76
	NH1 ASP A 325	OD1 ARG A 319	6.38
	NZ GLU A 331	OE1 LYS A 326	3.46
	NH2 GLU A 333	OE1 ARG A 335	6.43
	NH1 GLU A 336	OE1 ARG A 319	4.74
	NH1 GLU A 339	OE1 ARG A 362	5.03
	NH2 GLU A 349	OE1 HIS A 347	5.97
	ND1 ASP A 365	OD1 HIS A 378	3.06
	NE2 ASP A 375	OD1 HIS A 378	6.76
	NZ ASP A 382	OD1 HIS 346	4.42
	NE2 ASP A 400	OD1 HIS A 403	6.59
BCL-2	N ASP A 4	OD1 THR A 1	5.46
	NZ GLU A 7	OE1 LYS A 11	5.73
	NZ ASP A 25	OD2 LYS A 11	6.60
	NZ ASP A 96	OD1 LYS A 16	6.96
	NH1 ASP A 97	OD1 ARG A 100	5.66
	NH1 GLU A 130	OE1 ARG A 133	5.03
	NE ASP A 134	OD1 ARG A 133	6.54
	NH1 GLU A 154	OE1 ARG A 158	5.26
	NH1 GLU A 173	OE2 ARRG A 6	6.72

Prediction of Secondary Structure

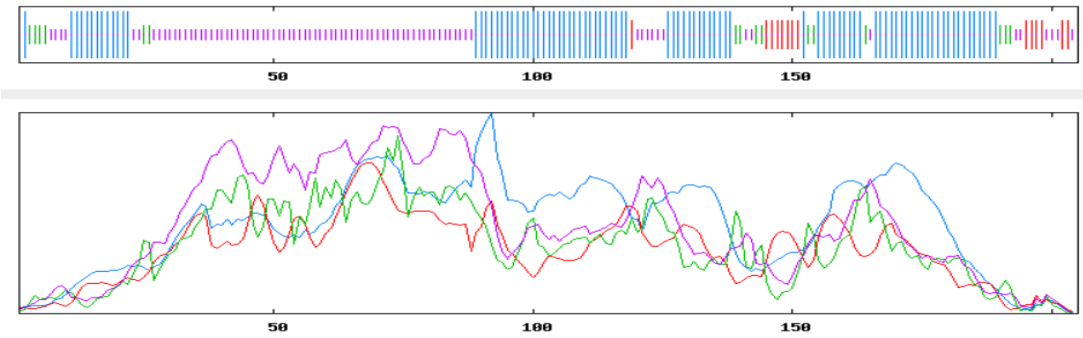
Figure 1 shows the SOPMA plots of p38, EGFR and Bcl-2 together with their respective alpha helix composition. In addition, the secondary structures of all three protein models are shown in Table 4 while their longest and shortest alpha helixes are shown in Table 5.



(A)



(B)



(C)

Figure 1: SOPMA plots for (A) p38, (B) EGFR and (C) Bcl-2

Table 4: Secondary structure of p38, EGFR and Bcl-2

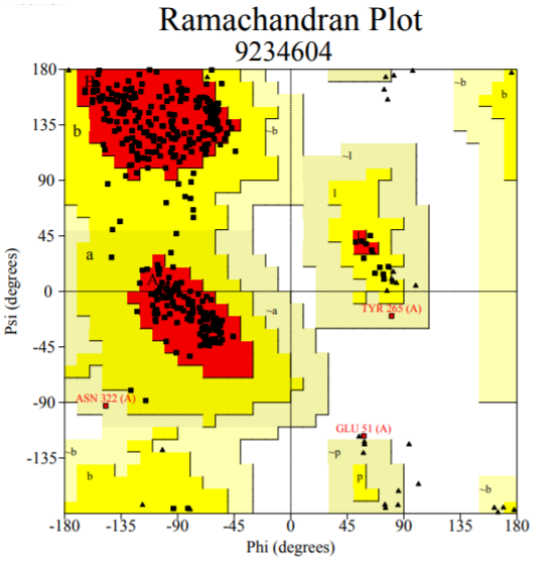
Secondary structure	Alpha-helix	Extended strand	Beta-turn	Random coil
p38	86	74	17	161
EGFR	81	76	20	287
Bcl-2	92	14	16	83

Table 5: Longest and shortest alpha helix of p38, EGFR and Bcl-2

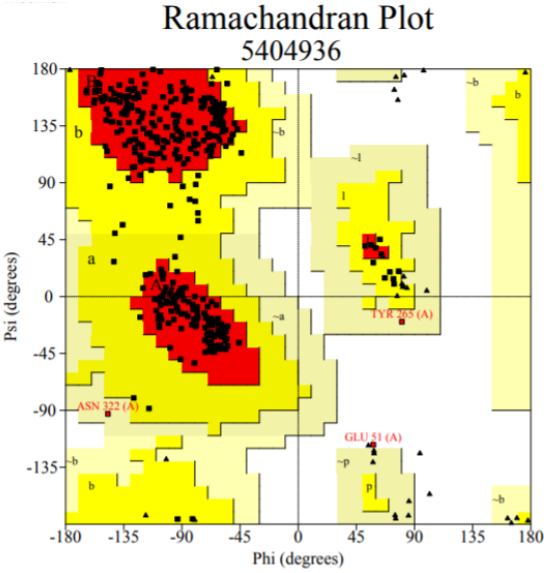
Amino acid	Longest alpha helix	Residues	Shortest alpha helix	Residues
p38	α-12	20	α-3, α-4, α-10	1
EGFR	α-6	15	α-2, α-4, α-7, α-9, α-12, α-17	1
Bcl-2	α-3	32	α-7, α-16, α-17, α-22, α-29, α-32, α-33, α-34, α-41, α-34, α-44	1

Validation of protein models

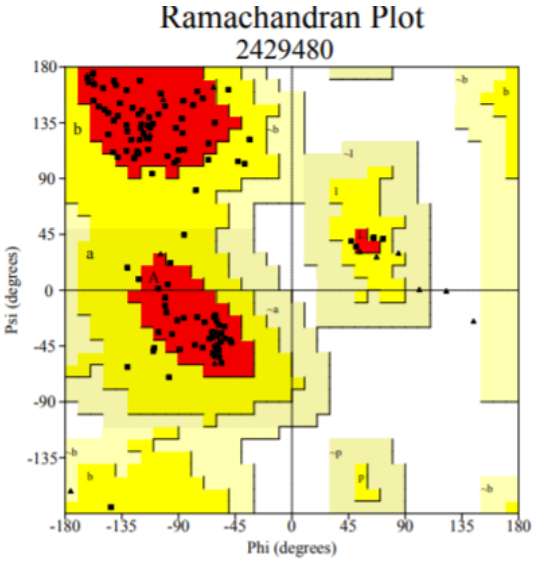
The stereochemistry and the residues of the three protein models were evaluated using the PROCHECK tool. Figure 2 shows the Ramachandran plot analysis of p38, EGFR and Bcl-2 proteins. Based on the analysis, it was confirmed that all three proteins are in the most favorable region which is above 80%. Table 6 shows the Ramachandran plot statistics of p38, EGFR and Bcl-2 proteins.



(A)



(B)



(C)

Figure 2: Ramachandran plot analysis for (A) p38, (B) EGFR and (C) Bcl-2proteins

Table 6: Validation of p38, EGFR and Bcl-2 proteins

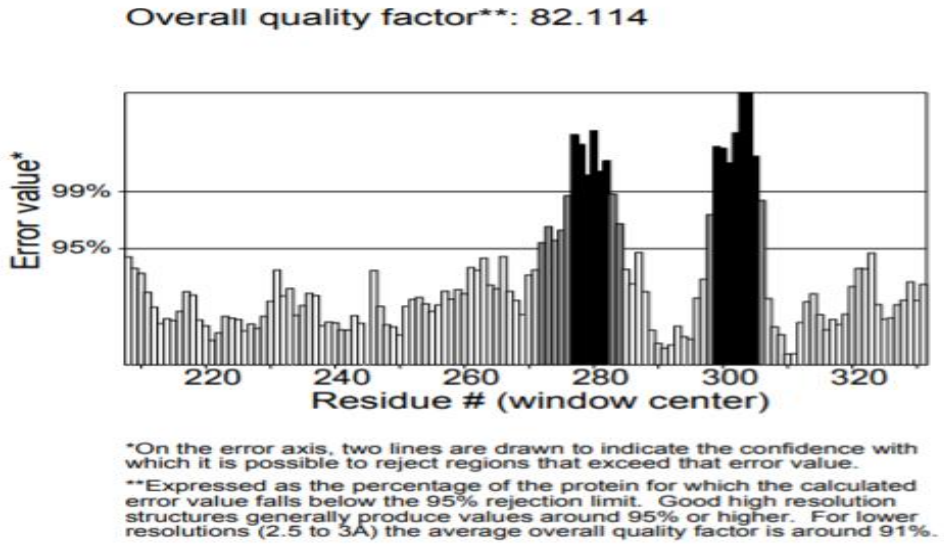
Structure	Ramachandran plot statistic (%)			
	Most favored	Additionally allowed	Generally allowed	Disallowed
p38	95	16	0	0
EGFR	90.6	31	2	1
Bcl-2	95	16	0	0

After that, ProQ was used to predict the quality of the protein models using Levitt-Gerstein (LG) and Maximum Subarray (MaxSub) score. Based on the results of LG score obtained from ProQ, p38, and Bcl-2 were considered to be good quality whereas EGFR had a very good quality. In terms of MaxSub score, all three proteins were of the correct quality. Table 8 shows the LG and MaxSub score of p53, EGFR and Bcl-2.

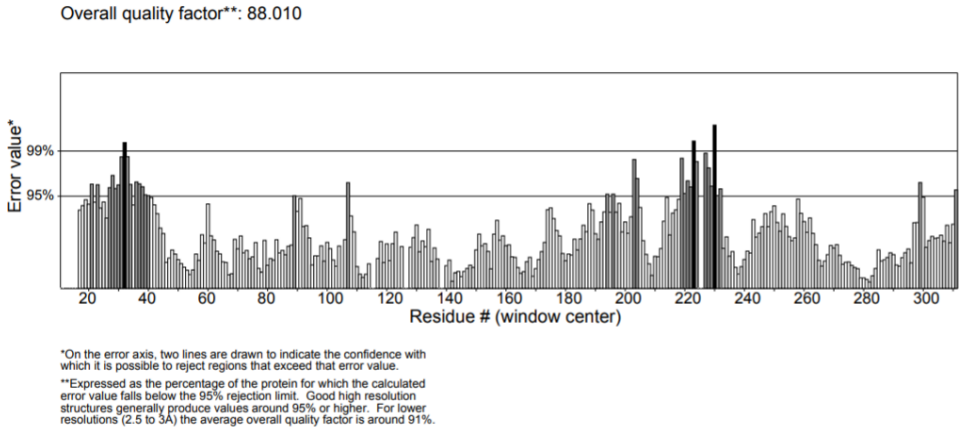
Table 7: LG and MaxSub score of p38, EGFR and Bcl-2

Structure	ProQ	
	LG score	**MaxSub score**
p38	4.927	0.586
EGFR	3.814	0.302
Bcl-2	4.927	0.586

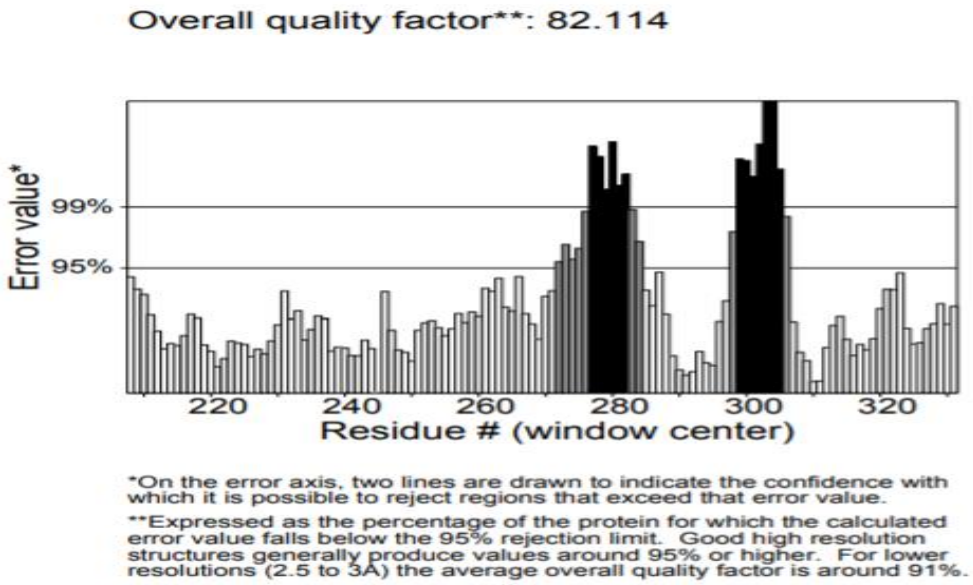
ERRAT is another protein validation tool using a crystallography technique in which its results refer to the overall quality factor of protein. A protein with a quality factor above 50% was predicted to be a good protein. According to the results of this study, the overall quality factor of p38, EGFR and Bcl-2 are 82.114%, 88.010% and 82.114% respectively. Thus, all three proteins are confirmed to be good quality. Figure 3 shows the ERRAT results of p38, EGFR and Bcl-2 proteins.



(A)



(B)



(C)

Figure 3: ERRAT results for (A) p38, (B) EGFR and (C) Bcl-2proteins

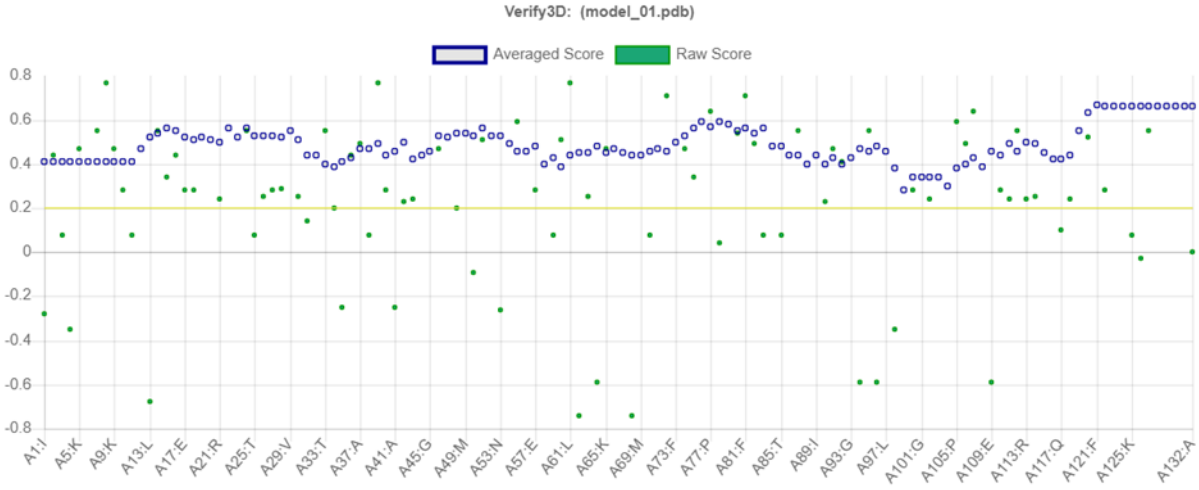
Verify 3D is another tool that is used to evaluate the quality of protein based on the number of residues available in each protein. Protein with residue amounts of more than 80% is predicted to be a good protein. According to the Verify 3D results, the number of residues of p38, EGFR and Bcl-2 was 100%, 98.59% and 100% respectively. Thus, all three proteins were confirmed to be of good quality. Figure 4 shows the Verify 3D result of p38, EGFR and Bcl-2.

Completed at 8:57 pm | View Structure

100.00% of the residues have averaged 3D-1D score >= 0.2

Pass

At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.



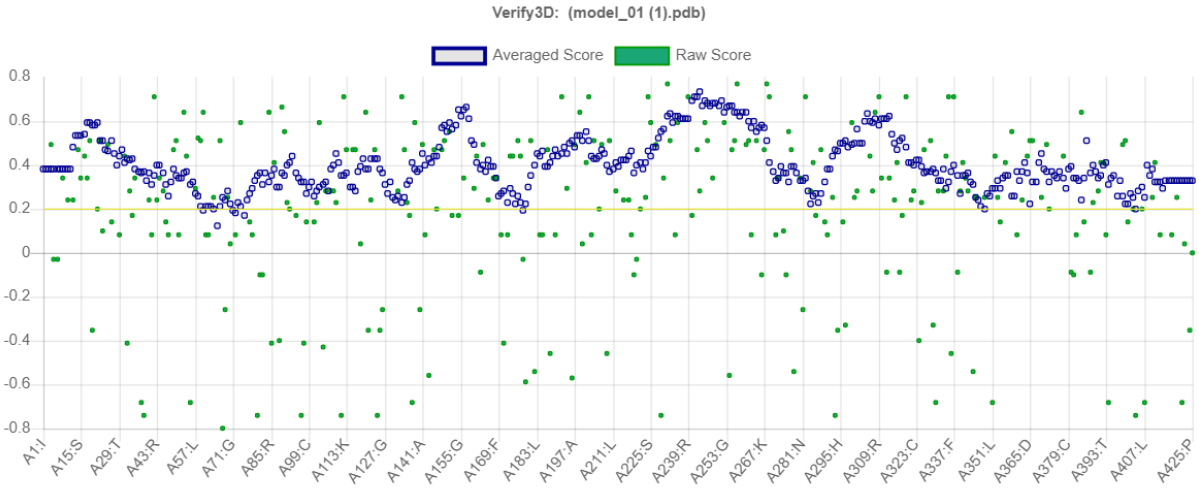
(A)

Completed at 9:03 pm | View Structure

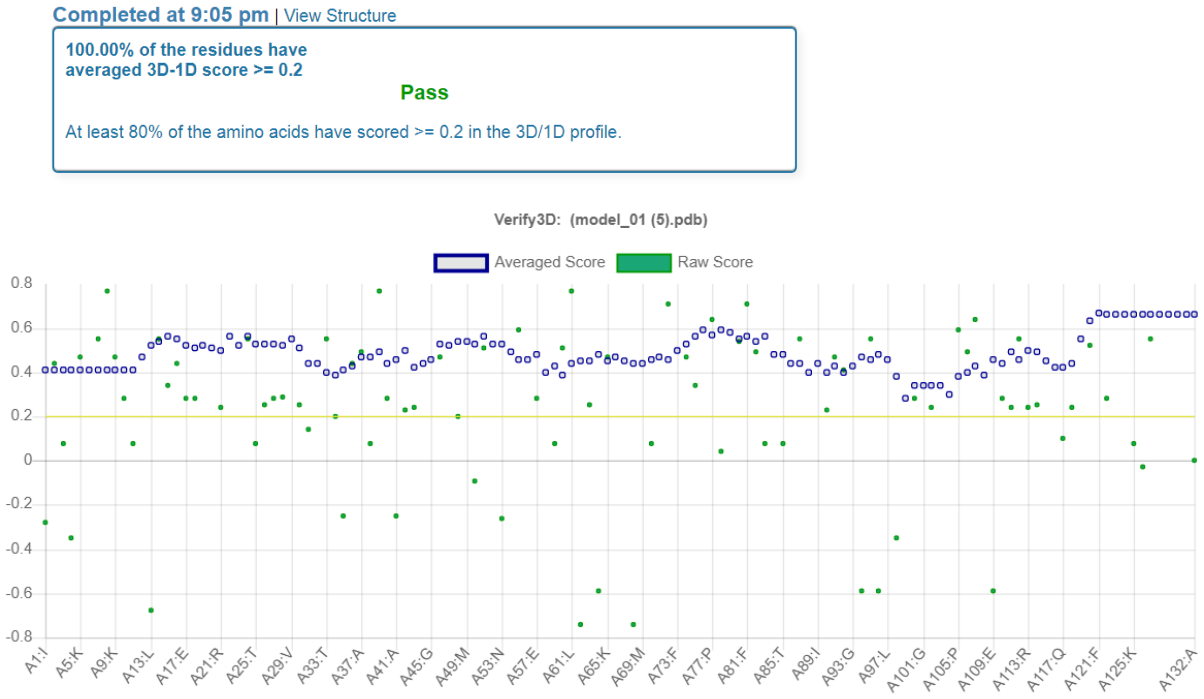
98.59% of the residues have averaged 3D-1D score >= 0.2

Pass

At least 80% of the amino acids have scored >= 0.2 in the 3D/1D profile.



(B)



(C)

Figure 4: Verify 3D results for (A) p38, (B) EGFR and (C) Bcl-2proteins.

Identification of Active Sites

The active sites of the three proteins were obtained using the active site-prediction server. Based on the results, the protein volume of p38, EGFR, Bcl-2 are 778 A^3, 1556 A^3 and 778 A^3 respectively. Table 8 shows the predicted active site of p38, EGFR and Bcl-2 proteins.

Table 8: Predicted active site of p38, EGFR, and BCL-2 proteins

Protein	Volume (A^3)	Pocket forming residues
p38	778	ILE1, GLU100, ARG102, GLY103, PRO105, PRO2, THR3, CYS4, GLY44, LYS46, LYS5, SER55, GLY56, GLU57, PHE58, THR59, ASP60, HIE66, VAL68, MET69, LYS70, ARG72, TRP76, GLU78, GLY79, HIE80, PHE81, THR83, THR85, MET99
EGFR	1556	CYS19, CYS20, HIE21, ASN22, GLN23, CYS24, ALA25, PRO31, ARG32, GLU33, SER34, ASP35, CYS36, VAL38, PRO384, ALA385, GLY386, VAL387, MET388, GLY389, CYS39, GLU390, ASN391, ASN392, THR393, LEU394, VAL395, TRP396, ARG40, LYS41, PRO410, ASN411, CYS412, THR413, TYR414, PHE42, ARG43,

		ASP44, GLU45, ALA46, THR47, CYS48, LYS49, ASP50, THR51, CYS52, PRO53, PRO54, LEU55, MET56, LEU57, TYR58, ASN68, GLU70, GLY71, LYS72, TYR73, SER74, GLY76, ALA77, THR78, CYS79, SER8, LYS81, GLY9, HIE92, GLY93
Bcl-2	778	ILE1, GLU100, ARG102, GLY103, PRO105, PRO2, THR2, THR3, CYS4, GLY44, LYS46, LYS5, SER55, GLY56, GLU57, PHE58, THR59, ASP60, HIE66, VAL68, MET69, LYS70, ARG72, TRP76, GLU78, GLT79, HIE80, PHE81, THR83, THR85, MET99

Docking of Target Protein-Plant Compound

In this study, BSP Slim Server was used to dock each target protein with its relevant phytocomponent. As a result, the highest docking score were chosen for each protein. Table 9 shows the best docking score of each protein which were obtained from BSP-Slim Server.

Table 9: Result of protein ligand docking using BSP Slim Server

Protein	Phytocomponent	Docking score
p38	Coumaric acid	-6.72
p38	Anthraquinone	-5.97
p38	Gingerol	-6.98
p38	Tanshinone	-6.52
p38	Baicalein	-7.70
p38	Wogonin	-7.21
EGFR	Coumaric acid	-7.38
EGFR	Anthraquinone	-6.53
EGFR	Gingerol	-7.41
EGFR	Tanshinone	-7.05
EGFR	Baicalein	-6.69
EGFR	Wogonin	-6.84
Bcl-2	Coumaric acid	-6.74
Bcl-2	Anthraquinone	-5.97
Bcl-2	Gingerol	-6.06
Bcl-2	Tanshinone	-5.41
Bcl-2	Baicalein	-6.54
Bcl-2	Wogonin	-5.13

Conclusion

In this study, each protein (p38, EGFR and Bcl-2) was successfully docked with its relative phytocomponent. The protein models had a strong interaction with the phytocomponents due to their low binding score. The best docking scores of the protein-phytocompound complexes (p38-wogonin, EGFR-gingerol, and Bcl-2-coumaric acid) were -7.21, -7.41, and -6.74 kcal/mol respectively.

REFERENCES

Angamuthu, K. and Piramanayagam, S., 2017. Evaluation of *in silico* protein secondary structure prediction methods by employing statistical techniques. *Biomedical and Biotechnology Research Journal (BBRJ)*, *1*(1), pp.29.

Appaiah, P. and Vasu, P., 2016. *In silico* designing of protein rich in large neutral amino acids using bovine αs1 Casein for treatment of Phenylketonuria. *Journal of Proteomics and Bioinformatics*, *9*(11), pp.287-297.

Beckles, M.A., Spiro, S.G., Colice, G.L. and Rudd, R.M., 2003. Initial evaluation of the patient with lung cancer: symptoms, signs, laboratory tests, and paraneoplastic syndromes. *Chest*, *123*(1), pp.97S-104S.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L. and Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(W1), pp.W252-W258.

Blair, A., Grauman, D.J., Lubin, J.H. and Fraumeni Jr, J.F., 1983. Lung cancer and other causes of death among licensed pesticide applicators. *Journal of the National Cancer Institute*, *71*(1), pp.31-37.

Collins, L.G., Haines, C., Perkel, R. and Enck, R.E., 2007. Lung cancer: diagnosis and management. *American Family Physician*, *75*(1), pp.56-63.

Costantini, S., Colonna, G. and Facchiano, A.M., 2008. ESBRI: a web server for evaluating salt bridges in proteins. *Bioinformation*, *3*(3), pp.137.

Crane, M., Scott, N., O'Hara, B.J., Aranda, S., Lafontaine, M., Stacey, I., Varlow, M. and Currow, D., 2016. Knowledge of the signs and symptoms and risk factors of lung cancer in Australia: mixed methods study. *BMC Public Health*, *16*, pp.1-12.

DeLano, W.L., 2004. Pymol reference guide. *Delano Scientific: San Carlos, CA, USA*, pp.1-68.

Elengoe, A. and Sebestian, E., 2020. *In silico* molecular modelling and docking of allicin, epigallocatechin-3-gallate and gingerol against colon cancer cell proteins. *Asia Pacific Journal of Molecular Biology and Biotechnology*, pp.51-67.

Kalet, I. J. 2014. Computational Models and Methods. *Principles of Biomedical Informatics* (Second Edition), pp. 479-578.

Kasilingam, T and Elengoe, A., 2018. *In silico* molecular modeling and docking of apigenin against the lung cancer cell proteins. *Asian Journal of Pharmaceutical and Clinical Research 11*(9), pp.246-252.

Łapińska, U., Saar, K.L., Yates, E.V., Herling, T.W., Müller, T., Challa, P.K., Dobson, C.M. and Knowles, T.P., 2017. Gradient-free determination of isoelectric points of proteins on chip. *Physical Chemistry Chemical Physics*, *19*(34), pp.23060-23067.

Mao, L., Hruban, R.H., Boyle, J.O., Tockman, M. and Sidransky, D., 1994. Detection of oncogene mutations in sputum precedes diagnosis of lung cancer. *Cancer Research*, *54*(7), pp.1634-1637.

Rami-Porta, R., Ball, D., Crowley, J., Giroux, D.J., Jett, J., Travis, W.D., Tsuboi, M., Vallieres, E., Goldstraw, P., Research, C. and International Staging Committee, 2007. The IASLC Lung Cancer Staging Project: proposals for the revision of the T descriptors in the forthcoming (seventh) edition of the TNM classification for lung cancer. *Journal of Thoracic Oncology*, *2*(7), pp.593-602.

Rigsby, R.E. and Parker, A.B., 2016. Using the P y MOL application to reinforce visual understanding of protein structure. *Biochemistry and Molecular Biology Education*, *44*(5), pp.433-437.

Sakiyama, T., Kohno, T., Mimaki, S., Ohta, T., Yanagitani, N., Sobue, T., Kunitoh, H., Saito, R., Shimizu, K., Hirama, C. and Kimura, J., 2005. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *International Journal of Cancer*, *114*(5), pp.730-737.

Singh, T., Biswas, D. and Jayaram, B., 2011. AADS-An automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. *Journal of Chemical Information and Modeling*, *51*(10), pp.2515-2527.

Supramaniam, G. and Elengoe, A., 2020. *In silico* molecular docking of glycyrrhizin and breast cancer cell line proteins. Plant-derived Bioactives: Chemistry and Mode of Action, pp.575-589.

Tiwari, A., Avashthi, H., Jha, R., Srivastava, A., Garg, V.K., Ramteke, P.W. and Kumar, A., 2016. Insights using the molecular model of Lipoxygenase from Finger millet (*Eleusine coracana* (L.)). *Bioinformation*, *12*(3), pp.156.

Verma, N.K. and Singh, B., 2013. Insight from the structural molecular model of cytidylate kinase from Mycobacterium tuberculosis. *Bioinformation*, *9*(13), pp.680.

Wallner, B. and Elofsson, A., 2003. Can correct protein models be identified?. *Protein Science*, *12*(5), pp.1073-1086.