1

In Silico Molecular Modeling and Docking Analysis in Lung Cancer Cell Proteins

Manisha Sritharan[1] and Asita Elengoe[2]*

[1]Department of Science and Biotechnology, Faculty of Engineering and Life Sciences, University of Selangor, Bestari Jaya, Selangor, Malaysia
[2]Department of Biotechnology, Faculty of Science, Lincoln University College, Petaling Jaya, Selangor, Malaysia

Abstract

In this study, the three-dimensional (3D) models of lung cancer cell line proteins [epidermal growth factor (EGFR), K-ras oncogene protein, and tumor suppressor (TP53)] were generated and their binding affinities with curcumin, ellagic acid, and quercetin through local docking were assessed. Firstly, Swiss model was used to build lung cancer cell line proteins and then visualized by the PyMol software. Next, ExPASy ProtParam Proteomics server was used to evaluate the physical and chemical parameters of the protein structures. Furthermore, the protein models were validated using PROCHECK, ProQ, ERRAT, and Verify3D programs. Lastly, the protein models were docked with curcumin, ellagic acid, and quercetin by using BSP-Slim server. All three protein models were adequate and in exceptional standard. The curcumin showed binding energy with EGFR, K-ras oncogene protein, and TP53 at 5.320, 2.730, and 1.633, kcal/mol, respectively. Besides that, the ellagic acid showed binding energy of EGFR, K-ras oncogene protein, and TP53 at −2.892, 0.921, and 0.054 kcal/mol, respectively. Moreover, the quercetin showed binding energy of EGFR, K-ras oncogene protein, and TP53 at −1.249, −1.154, and −0.809 kcal/mol, respectively. The EGFR had the strongest bond with ellagic acid while K-ras oncogene protein and TP53 had the strongest interaction with quercetin. In order to identify the appropriate function, all these potential drug candidates can be further assessed through laboratory experiments.

***Keywords*:** EGFR, K-ras, TP53, curcumin, ellagic acid, quercetin, docking

**Corresponding author*: asitaelengoe@yahoo.com

R. J. Hemalatha, D. Akila, D. Balaganesh and Anand Paul (eds.) The Internet of Medical Things (IoMT): Healthcare Transformation, (1–22) © 2022 Scrivener Publishing LLC

1.1 Introduction

Lung cancer is known to be the number one cause of cancer deaths among all the cancer in both men and women in worldwide. According to a World Health Organization (WHO) survey, lung cancer caused 19.1 deaths per 100,000 in Malaysia, or 4,088 deaths per year (3.22% of all deaths) [1]. Moreover, there was a record of 1.69 million of deaths worldwide in 2015 due to lung cancer. Furthermore, a research in UK estimated that there will be 23.6 million of new cases of cancer worldwide each year by 2030 [1]. The main cause of lung cancer deaths is smoking. Almost 8% of people died because of it. Furthermore, the second reason is exposure to second-hand smoke. Thus, it is very clear that smoking is the leading risk factor for lung cancer. However, not everyone who got lung cancer is smokers as many people with lung cancer are former smokers while many others never smoked at all. Moreover, radiation exposure, unhealthy lifestyle, secondhand smoke, pollution of air, genetic markers, prolongs inhalation of asbestos, and chemicals as well as other factors can cause lung cancer non-smokers [2].

Furthermore, it seems that most lung cancer signs do not appear until the cancer has spread, although some people with early lung cancer do have symptoms. Generally, the symptoms of lung cancer are a cough that does not go away and instead gets worse, shortness of breath, chest pain, feeling tired or weak, new onset of wheezing, and some lung cancer can even cause syndrome [3]. On top of that, a number of tests can be conducted in order to look for cancerous cell such as X-ray image of lung that could disclose the abnormal mass or nodule, a CT scan to exhibit small lesions in the lungs which may not detected on X-ray, blood investigations, sputum cytology, and tissue biopsy [4]. Lung cancer treatments being carried out are adjuvant therapy which may include radiation, chemotherapy, targeted therapy, or immunotherapy.

Since they originate from the bronchi within the lungs, small-cell lung carcinoma (SCLC) and non–small-cell lung carcinoma (NSCLC) are the two main clinic pathological classes for lung cancer. They are also known as bronchogenic carcinomas because they arise from the bronchi within the lungs [4]. Lung cancer is believed to be arising after a series of continuous pathological changes (preneoplastic lesions) which very often discovered accompanying lung cancers as well as in the respiratory mucosa of smokers. Apart from that, the genes involved in lung cancer are epidermal growth factor receptor (EGFR), KRAS, MET, LKBI, BREF, ALK, RET, and tumor suppressor gene (TP53) [5].

The three most common genes in lung cancer are EGFR, KRAS, and TP53, and the structure of these genes is explored in thus study. EGFR is actually transmembrane protein that has cytoplasmic kinase movement and it transduces essential development factor motioning from the extracellular milieu to the cell. According to da Cunha Santos, more than 60% of NSCLCs expresses EGFR which has turned into an essential focus for the treatment of these tumors [6]. In addition, the KRAS mutation is the most widely recognized oncogene driver change in patients with NSCLC and presents a poor guess in the metastatic setting, making it an imperative focus for tranquilize advancement. It is difficult to treat patients with KRAS mutations since there is no targeted therapy yet [7]. Among the mutations, the most common mutation that found to occur in lung cancer is TP53 mutations and its frequency becomes greater with tobacco consumption [8]. In this study, three compounds (curcumin, ellagic acid, and quercetin) were used for docking with the three mutant proteins. Curcumin has excellent safety profile that focus on different infections with solid confirmation on molecule level. Thus, improvement in formulation criteria can aid in developing therapeutic drug [9]. Next, ellagic acid has the ability to bind with cancer cells to make them inactive and it is also effective to resist cancer in rats and mice according to a research [10]. Quercetin is a pigment from plant (flavonoid) which has anti-oxidant and anti-inflammatory effect. It has shown to inhibit the multiplication of cancer cells according to Pao-Chen Kuo *et al*. [11].

Bioinformatics is a multidisciplinary discipline that creates methods and software tools for storing, extracting, organizing, and interpreting biological data. To analyze biological data, a combination of bioinformatics and computer science, statistics, physics, chemistry, mathematics, and engineering is useful. Currently, this method is growing rapidly because it is cheap and quite faster than experimental approaches. Computational biology tools such as protein modeling (e.g., Swiss Model, Easy Modeller, and Modeller), molecular dynamic simulation (e.g., Gromacs and Amber), and docking (e.g., Autodock version 4.2, AutodockVina, Swissdock, and Haddock) helped design substrate-based drugs to study the interaction between the target proteins (cancer cell proteins) and ligand (phytocomponents).

The aim of conducting this research is to initiate three-dimensional (3D) models of lung cancer line proteins (EGFR, K-ras oncogene, and TP53) and to guesstimate their binding affinities with curcumin, ellagic acid, and quercetin via docking approach.

1.2 Methodology

1.2.1 Sequence of Protein

The entire amino acid sequence of the EGFR (GI: 110002567), K-ras oncogene protein (GI: 186764), and TP53 (GI: 1233272225) were obtained from National Center for Biotechnology Information Center for Biotechnology Information (NCBI). Next, EGFR consists of 464 amino acids, K-ras oncogene protein contains 188 amino acids, while TP53 consists of 346 amino acids.

1.2.2 Homology Modeling

As of now, the 3D models of EGFR, K-ras oncogene protein, and TP53 are not available in Protein Data Bank (PDB). As a result, the models were started with Swiss Model [12] and then visualized with PyMol [13].

1.2.3 Physiochemical Characterization

The physical and chemical characters of the protein structures were analyzed by using the ExPASy ProtParam Proteomics tool [14]. Besides that, hydrophobic and hydrophilic were foreseen by ColorSeq analysis [15]. Furthermore, The ESBRI program [16] was used to reveal salt bridges in protein structures, while the Cys Rec program was used to count the number of disulfide bonds [17].

1.2.4 Determination of Secondary Models

The secondary structural properties were discovered using the Alignment Self-Optimized Prediction Process (SOPMA) [18].

1.2.5 Determination of Stability of Protein Structures

PROCHECK was used to test the protein models [19]. ProQ [20], ERRAT [21], and Verify3D programs were used to conduct further research [22].

1.2.6 Identification of Active Site

The 3D model of EGFR, K-ras oncogene protein, and TP53 were submitted to active site-prediction server [23] in order to discover their binding sites.

1.2.7 Preparation of Ligand Model

The tertiary structure of the quercetin, curcumin, and ellagic acid are not openly accessible. The whole sequence of quercetin, curcumin, and ellagic acid were attained from PubChem, National Center for Biotechnology Information (2017) [24].

1.2.8 Docking of Target Protein and Phytocompound

The 3D structure of EGFR was docked with quercetin, curcumin, and ellagic acid by using BSP-Slim server [25]. In addition, the best docking complex model was chosen based on the lowest binding score. The similar docking method was carried out between the other two protein models and phytocompounds (quercetin, curcumin, and ellagic acid).

1.3 Results and Discussion

1.3.1 Determination of Physiochemical Characters

The isoelectric point (pI) value quantified for EFGR (pI > 7) specify basic feature while the pI for k-ras and TP53 (pI < 7) exhibit acidic. Besides that, the molecular weight of EFGR, k-ras oncogene protein, and TP53 are 50,343.70, 21,470.62, and 38,532.60 Daltons, respectively. The extent of light being by absorbed by protein at a specific wavelength was used to calculate the extinction coefficient of TYR, TRP, and CYS residues where for EGFR is 38,305 M/cm, k-ras oncogene protein is 12,170M/cm, and TP53 is 43,025 M/cm. In addition, −R is the negatively (ASP + GLU) and +R is the positively charged (ARG + LYS) residues in the amino acid sequence. The total of −R and +R for each protein model was described in Table 1.1. According to the instability index of ExPASy ProtParam, EGFR proteins are classified as stable because the instability index for both proteins are less than 40 while K-ras oncogene protein and TP53 is categorized as unstable as the instability index is more than 40. The instability index for EGFR is 35.56, K-ras oncogene protein is 43.95, and TP53 is 80.17. On top of that, the very low grand average of hydropathicity (GRAVY) index (a (negative value GRAVY) of EGFR, K-ras oncogene protein, and TP53 denotes their hydrophilic nature (Table 1.1). Apart from that, EFGR, K-ras and TP53 have more polar residues (41.52%, 53.33%, and 45.29%) than non-polar residues (26.74%, 30.0%, and 27.35%) which were determined using Color Protein Sequence.

Table 1.1 Physiochemical characters of EGFR, K-ras, and TP53 proteins as determined by ExPASy ProtParam program.

Protein	Length	Molecular weight (kDa)	pI	−R	+R	Extinction coefficient	Instability index	Aliphatic index	GRAVY
EGFR	460	50,343.70	7.10	49	49	38,305	35.56	72.91	−0.269
KRAS	180	21,470.62	8.18	29	31	12,170	43.95	77.18	−0.559
TP53	340	38,532.60	5.64	41	33	43,025	80.17	63.99	−0.592

Furthermore, the structure and function of the protein can be affected by salt bridges. Thus, salt bridge disruption minimizes the stability of protein [26]. Next, it is also associated with regulation, molecular recognition, oligomerization, flexibility, domain motions, and thermostability [27, 28].

The greater number of arginine in the protein model enhances the stability of the protein. This is happens through the electrostatic interactions between their guanidine group [29]. Hence, it was confirmed that all the protein models are in the identical stable conditions. The outcome of Cys_Recserver exhibits that the quantity of disulfide bonds in EGFR is 42, K-ras oncogene protein is 5, and TP53 is 11 (Table 1.2).

1.3.2 Prediction of Secondary Structures

Results from SOPMA analysis shows that random coils dominant among secondary structure components in the protein models (Figure 1.1). The constitution of alpha helix in EGFR, K-ras oncogene protein, and TP53 were shown in Table 1.3.

The outcome from this analysis specified that EGFR, K-ras oncogene protein, and TP53 constitutes of 15, 11, and 10α helices, respectively. Besides that, Table 1.4 represents the details of the longest and shortest alpha helix of all the protein models.

1.3.3 Verification of Stability of Protein Structures

PROCHECK server was used to verify the stereo chemical quality and the geometry of protein models through Ramachandran plots (Figure 1.2). Furthermore, it was revealed that all the protein structures are in most favorable region because they had percentage value more than 80% (Table 1.5). Thus, the standard of these proteins was assessed to be immense and reliable. On top of that, PROCHECK analysis disclose that a number of residues such as TYR265 and GLU51 for EGFR while LYS180 for K-ras oncogene protein were located away from energetically favored regions of Ramachandran plot. Besides that, there are no residues found at forbade region for TP53 protein model.

Thereby, the stereo chemical interpretation of backbone phi/psi dihedral angles deduced that EGFR, K-ras oncogene protein, and TP53 have low percentage of residues among the protein models. Moreover, ProQ was utilized in order to validate "the quality" with the usage of Levitt-Gerstein (LG) score and maximum subarray (MaxSub). All the protein models were within the range for LG and MaxSub score according to the outcome exhibited for creating a good model (Table 1.5).

Table 1.2 The number disulfide bonds were quantitated by Cys_Rec prediction program.

Protein	Cys_Rec	Score
EGFR	Cys_9	−13.0
	Cys_13	39.2
	Cys_17	100.1
	Cys_25	98.3
	Cys_26	104.2
	Cys_30	104.1
	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

(*Continued*)

Table 1.2 The number disulfide bonds were quantitated by Cys_Rec prediction program. (*Continued*)

Protein	Cys_Rec	Score
	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_365	70.2
	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_430	3.7
KRAS	Cys_12	−28.5
	Cys_51	−74.2
	Cys_80	−72.6
	Cys_118	−56.4
	Cys_185	−15.2

(*Continued*)

Table 1.2 The number disulfide bonds were quantitated by Cys_Rec prediction program. (*Continued*)

Protein	Cys_Rec	Score
TP53	Cys_124	−19.4
	Cys_135	−1.6
	Cys_141	−17.9
	Cys_176	−9.1
	Cys_182	−45.1
	Cys_229	−54.4
	Cys_238	1.6
	Cys_242	−5.5
	Cys_275	−34.4
	Cys_277	−51.5
	Cys_339	−32.8

ERRAT analysis is used for assessing the protein models which were determined by x-ray crystallography. Next, the value of ERRAT relies upon the statistics of non-bonded atomic interactions in the 3D protein structures. The protein is generally accepted as high quality protein if the percentage is greater than 50%. The ERRAT analysis score result shows that K-ras oncogene protein had the highest at 94.767. Therefore, it can be seen that K-ras oncogene protein has high quality resolution among the protein models. Besides that, the score value for EGFR is 88.010 while 90.374 for TP53 (Figure 1.3).

The Verify3D server was used to reveal the residues in each protein in which EGFR, K-ras oncogene, and TP53 had 98.59%, 100.00%, and 92.96% residues, respectively. Next, the average 3D-1D score of all three proteins are more than 0.2. As a consequence, it specifies that all of the sequences were in line with its protein model (Figure 1.4). Certainly, the resulting energy minimized EGFR, K-ras oncogene protein, and TP53 protein models satisfied the standard for evaluation of protein. Hence, the docking analysis with ligand will be carried out.

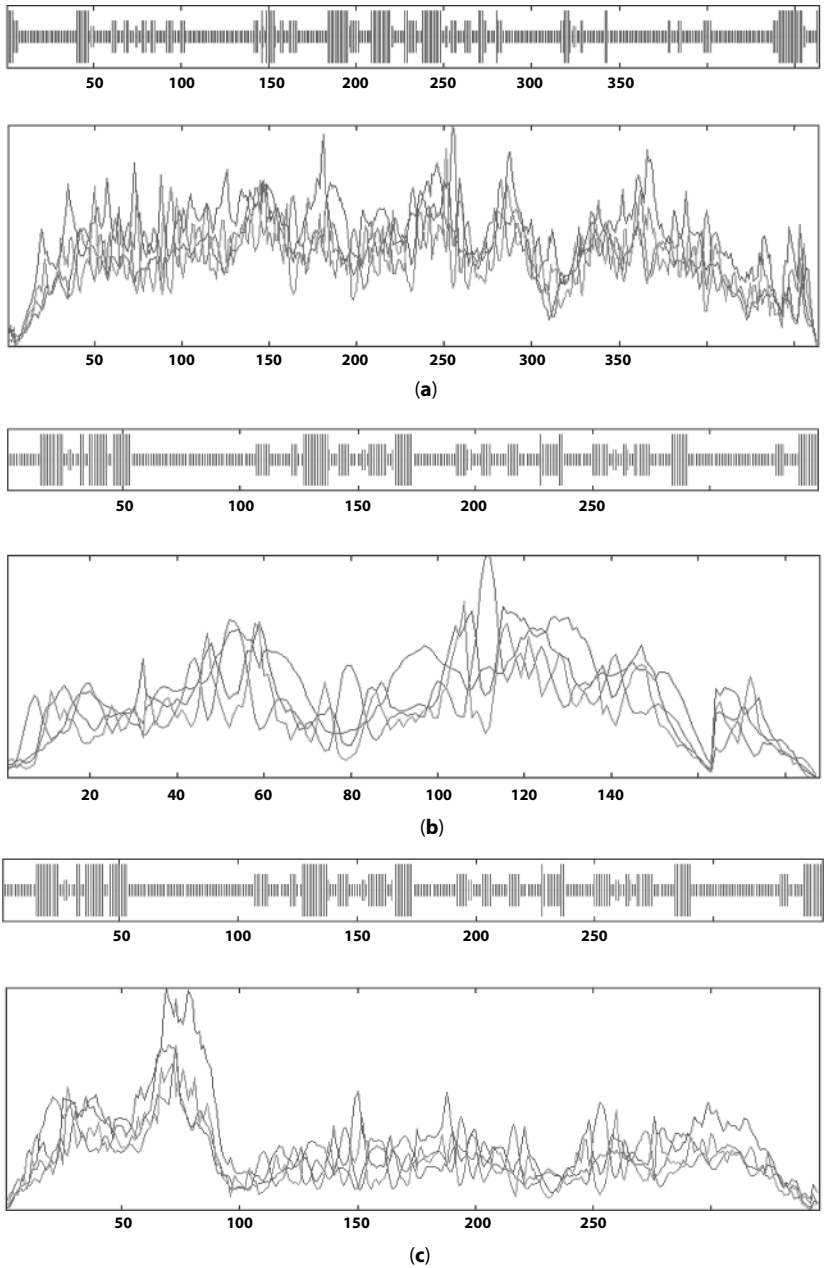

Figure 1.1 SOPMA plots for (a) EGFR, (b) K-ras oncogene protein, and (c) TP53.

Table 1.3 Secondary structure of the EGFR, K-ras oncogene protein, and TP53.

Secondary structure	Alpha helix (Hh)	Extended strand (Ee)	Beta turn (Tt)	Random coil (Cc)
EGFR	16.81	16.81	3.23	64.44
KRAS	43.62	21.81	7.45	27.13
TP53	18.79	18.21	3.18	59.83

Table 1.4 Composition of α-helix EGFR, K-ras oncogene protein, and TP53.

Amino acid	Longest alpha helix	Residues	Shortest alpha helix	Number of residues
EGFR	α_{14}	14	α_3, α_6, α_{11}, α_{15}	1
KRAS	α_{11}	20	α_1, α_{10}	1
TP53	α_5	11	α_7	1

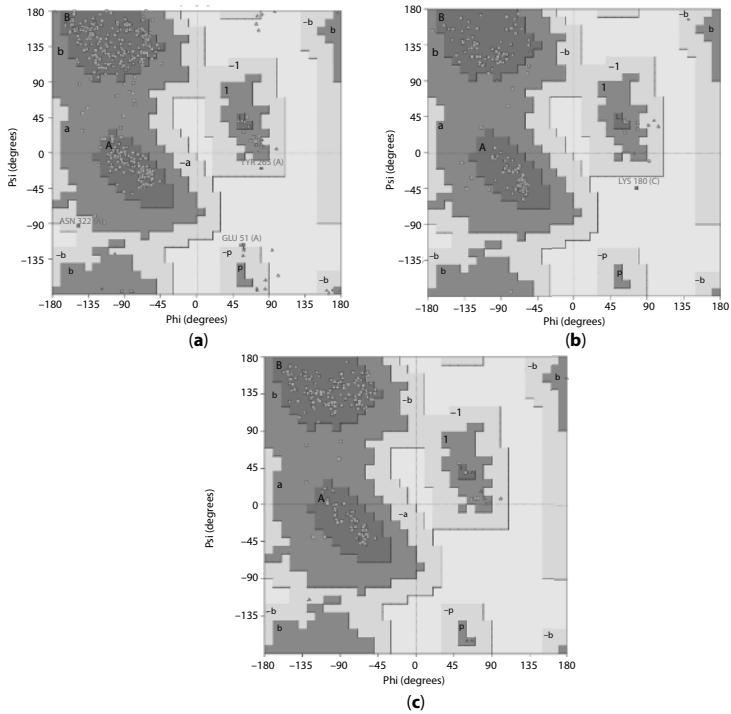

Figure 1.2 Ramachandran plots for (a) EGFR, (b) K-ras oncogene protein, and (c) TP53.

Table 1.5 Validation of the EGFR, K-ras oncogene protein, and TP53.

	Ramachandran plot statistics				Goodness factor			ProQ	
Structure	Most favored	Additionally allowed	Generously allowed	Disallowed	Dihedral angles	Covalent forces	Overall average	LG score	MaxSub
EGFR	90.6	8.6	0.6	0.3	−0.27	0.02	−0.14	3.814	0.302
KRAS	90.5	8.9	0.0	0.6	−0.12	0.04	−0.04	4.094	0.474
TP53	92.9	7.1	0.0	0.0	−0.16	0.07	−0.06	4.417	0.454

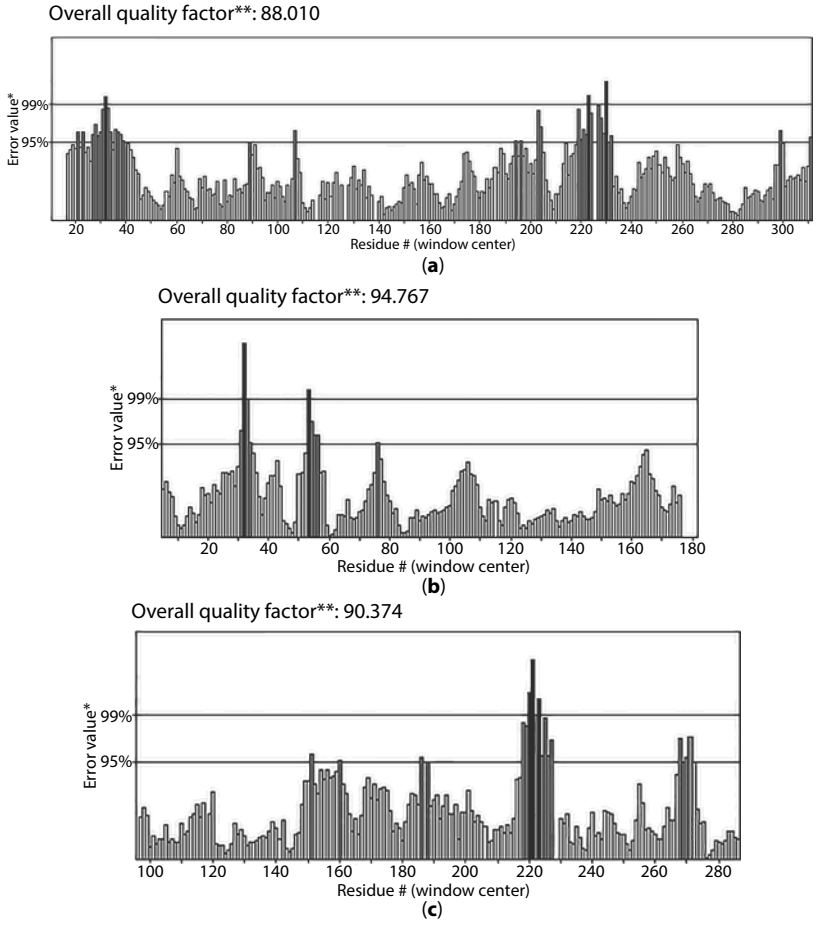

Figure 1.3 ERRAT plots for (a) EGFR, (b) K-ras oncogene protein, and (c) TP53.

1.3.4 Identification of Active Sites

For EGFR, K-ras oncogene protein, and TP53, the BSP-Slim server was used to obtain the active site protein volume and the residues that form an active site pocket (Table 1.6). The protein volume for EGFR, K-ras, and TP53 were 837 A^3, 718A^3, and 647A^3, respectively.

1.3.5 Target Protein-Ligand Docking

Based on Murugesan *et al.* study [30], the plant compounds from methanolic leaf extract of *Vitexnegundo*weredocked successfully with cyclooxygenase-2 (COX-2) enzyme. The phytocompounds had a better interaction

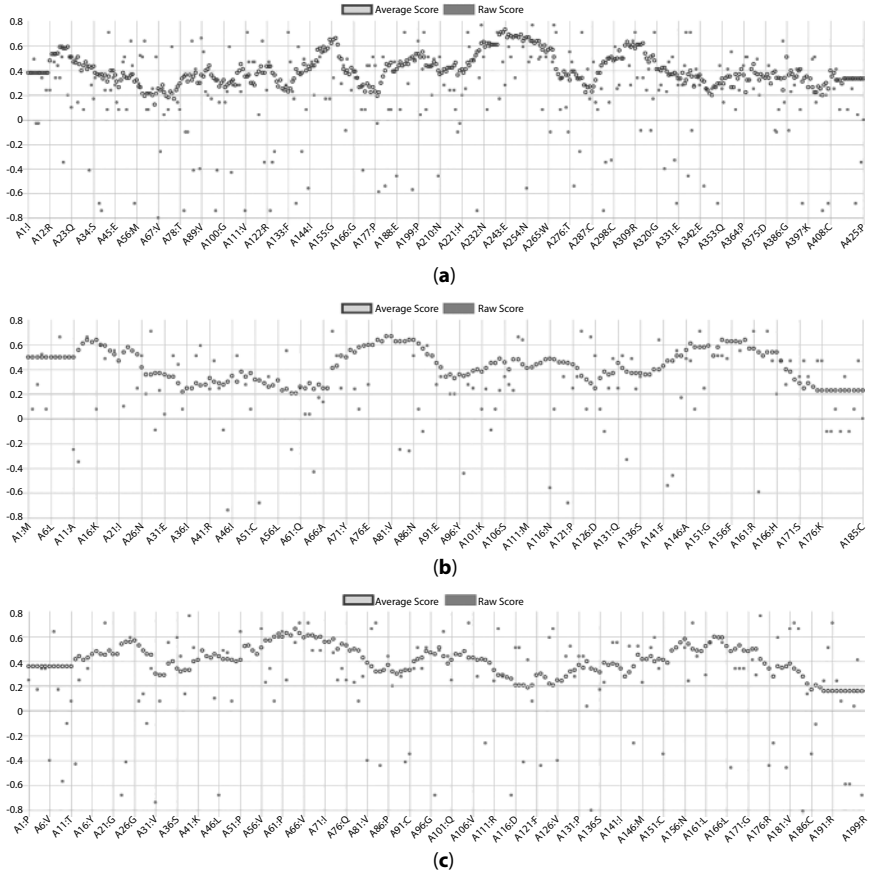

Figure 1.4 Verify 3D plots for (a) EGFR, (b) K-ras oncogene protein, and (c) TP53.

compared with aspirin and ibuprofen. They had a good binding energy and docking result.

Besides that, four components [1,3-Dioxolane, 2-(3-bromo-5,5,5-trichloro-2,2-dimethylpentyl)-, Butanoic acid, 2-hydroxy-2-methyl-methyl ester, DL-3,4-Dimethyl-3,4-hexanediol, and Pantolactone] from *Moringaconcanensis*had good binding affinity with brain cancer receptors. The binding energies were −3.90, −2.75, −3.05, and −4.15 kcal/mol. They had the lowest binding energies [31].

According to Deepa *et al.* study, plant compounds from the ethanolic leaf extract of *VitexNegundo* [(4S)-2-Methyl-2-phenylpentane-1,4-diol, 7-Methoxy-2,3-dihydro-2-phenyl-4-quinolone, 3-(tert-Butoxycarbonyl)-6-(3-benzoylprop-2-yl)phenol, (3R,4S)-4-(methylamino)-1-phenylpent-1-en-3-ol, and (2S,1'S)-1-Benzyl-2-[1'-(dibenzylamino) ethyl]aziridine] were docked

Table 1.6 Predicted active sites of the EGFR, K-ras oncogene protein, and TP53.

Protein	Volume	Residues that forming pocket
EGFR	837	GLY100, ALA 101, ASP102, SER103, TYR104, GLU105, MET106, GLU107, GLU108, LYS113, LYS115, LYS 116, CYS117, GLU118, GLY119, PRO120, CYS121, ARG122, LYS123, VAL124, ASN149, THR151, SER152, SER154, THR185, LYS187, GLU188, THR190, ASN210, GLU212, ILE213, ARG215, LYS242, CYS99
K-ras oncogene protein	718	GLY10, ALA11, CYS12, GLY13, VAL14, GLY15, LYS16, ASP33, PRO34, THR35, GLU37, LEU56, ASP57, THR58, ALA59, GLY60, GLN61, GLU62, GLU63, SER65, ARG68, MET72, ALA83, ASN86, LYS88, SER89, GLU91, ASP92, ILE93, HIE94, HIE95, TYR96, ARG97, GLU98, GLN99, ILE100, ARG102, VAL103
TP53	647	GLU107, ASN109, THR11, PRO128, TYR129, GLN13, GLU130, PRO131, PRO132, GLU133, VAL134, GLY135, SER136, ASP137, CYS138, THR139, THR140, ILE141, HIE142, TYR143, TYR16, GLY17, ASN177, SER178, PHE18, ARG19, LEU20, GLY21, PHE22, LEU23, HIE24, TYR35, ASN40, MET42, THR49, CYS50, PRO51, GLN53, LEU54, TRP55, VAL56, ASP57, THR59, PRO60, PRO61, THR64

with glucosamine 6 phosphatase synthase. They had the lowest and most negative value for binding energy (−36.53, −33.57, −35.90, −33.88, and −37.65 kcal/mol) [32].

According to Kasilingam and Elengoe study, apigenin successfully docked with p53, caspase-3, and MADCAM1 using BSP-Slim server. Apigenin was the plant compound while p53, caspase-3, and MADCAM1 were the target proteins in lung cancer cell line. Apigenin bound strongly with p53, caspase-3, and MADCAM1 at the lowest binding energies (4.611, 5.750, and 5.307 kcal/mol, respectively) [33].

Based on Ashwini *et al.* study, coumarin, camptothecin, epigallocatechin, quercetin, and gallic acid were screened for potential binding with

caspase-3 (target protein) in human cervical cancer cell line (HeLa). Coumarin had the strongest interaction with caspase-3 at the lowest binding affinity (−378.3 kJ/mol). Therefore, it could be a potential anti-cancer drug. However, gallic acid had the least interaction with caspase-3 at the lowest binding energy (−181.3 kJ/mol). The docking approach was carried out using Hex 8.0.0 docking software [34].

Chakrabarty *et al*. study demonstrated that 1-hexanol and 1-octen-3-ol suppressed the enzyme activity of Ach (PDB id: 2CKM) and BACE1 (PDB id: 4IVT). Ach and BACE1 are the proteins responsible for Alzheimer disease. 1-hexanol and 1-octen-3-ol were the plant compounds derived from leaf extract of *Lantana Camera (L.).* Glide Standard Precision (SP) ligand docking was performed to determine the binding energy. The results show that 1-hexanol and 1-octen-3-ol bound strongly with Ach at −2.291 and −2.465 kJ/mol, respectively. Whereas, 1-hexanol and 1-octen-3-ol had the lowest binding affinity with BACE 1 at −0.948 and −1.267 kJ/mol, respectively. 1-octen-3-ol may have the potential to be an effective drug against Alzhemeir disease. It had the best interaction with both enzymes (Ach and BACE1) when compared with 1-hexanol [35].

Based on Supramaniam and Elengoe study, glycyrrhizin successfully docked with p53, NF-kB-p105, and MADCAM1 using BSP-Slim server. Glycyrrhizin was the plant compound while p53, NF-kB-p105, and MADCAM1 were the target proteins in breast cancer cell line. Glycyrrhizin bound strongly with p53, NF-kB-p105, and MADCAM1 at the lowest binding affinities (−4.040, −5.127, and −5.251 kcal/mol, respectively). Therefore, glycyrrhizin could be a potential drug in breast cancer treatment [36].

According to Elengoe and Sebestian study, p53, adenomatous polyposis coli (APC), and EGFR were generated using homology modeling approach. These proteins were the target proteins. They were docked successfully with plant compounds such as allicin, epigallocatechin-3-gallate, and gingerol. Plant compounds were used as ligands in docking process. p53 had the most stable interaction with the allicin among the three target proteins. p53 docked with allicin at the lowest binding energy of 4.968. However, the other target proteins had the good docking score too [37].

In this study, EGFR is successfully docked with quercetin, curcumin, and ellagic acid by using the BSP-Slim server. The same target protein-phytocompound complex docking method was repeated with K-ras oncogene protein and TP53. Furthermore, the most suitable docking complex was selected based on the lowest binding energy ($D_G$bind). Results of docking showed that EGFR had a strong bond with ellagic acid since it was the most favorable with the lowest energy value (−2.892 kcal/mol) when compared to curcumin and quercetin (Table 1.7). In addition, there was

Table 1.7 Docking result of the EGFR, K-ras oncogene protein, and TP53.

Protein	Compounds	Binding energy (kcal/mol)
EGFR	Curcumin	5.320
	Ellagic acid	−2.892
	Quercetin	−1.249
K-ras oncogene protein	Curcumin	2.730
	Ellagic acid	0.921
	Quercetin	−1.154
TP53	Curcumin	1.633
	Ellagic acid	0.054
	Quercetin	−0.809

strong interaction between K-ras oncogene protein and quercetin with lowest energy (−1.154 kcal/mol) that was most favorable when compared to curcumin and ellagic acid. In addition, the strongest interaction for TP53 was with quercetin when compared to other two compounds with lowest energy (0.809 kcal/mol) according to the docking analysis.

1.4 Conclusion

In a nutshell, EGFR was successfully docked with curcumin, ellagic acid, and quercetin. Besides that, the same approach of docking simulation was performed for K-ras oncogene protein and TP53. Among the three protein models, EGFR had a strong interaction with ellagic acid due to the lowest energy value while K-ras oncogene protein and TP53 had a strong interaction with quercetin as the binding energy was the lowest. Consequently, result from this study will aid in designing a suitable structure-based drug. However, wet lab must be carried out to verify the results of this study.

References

1. Cancer Research UK, *Worldwide cancer statistics*, 2012, https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer#collapseZero.

2. American Cancer Society, *Lung cancer prevention and early detection*, 2016, https://www.cancer.org/cancer/lung-cancer/prevention-and-early-detection/signs-and-symptoms.html.
3. American Cancer Society, *Causes, risk factors and prevention*, 2016, https://www.cancer.org/cancer/non-small-cell-lung-cancer/causes-risks-prevention/what causes.
4. Mayo Clinic, *Lung cancer*, 2018, https://www.mayoclinic.org/diseases-conditions/lung-cancer/diagnosis-treatment/drc-20374627.
5. El-Telbany, A. and Patrick, C.M., Cancer genes in lung cancer. *Genes Cancer*, M3, 7–8, 467–480, 2012.
6. Santos, D.C., Sheperd, F.A., Tsao, M.S., EGFR mutations and lung cancer. *Annu. Rev. Pathol.: Mechanisms of Disease,* 6, 49–69, 2016.
7. Bhattacharya, S., Socinski, M.A., T.F., KRAS mutant lung cancer: progress thus far on an elusive therapeutic target. *Clin. Transl. Med.*, 4, 35, 2015.
8. Halverson, A.R., Silwal-Pandit, L., Meza-Zepeda, L.A. *et al.*, TP53 mutation spectrum in smokers and never smoking lung cancer patients. *Front. Genet.*, 7, 85, 2016.
9. Basnet, P. and Skalko-Basnet, N., Curcumin: An anti-inflammatory molecule from a curry spice on the path to cancer treatment. *Molecules*, 6, 6, 4567–4598, 2011.
10. Healthline, Whyellagic acid is important?, 2020, https://www.healthline.com/health/ellagic-acid.
11. Kuo, P.-C., Liu, H.-F., Chao, J.-I., Survivin and p53 modulate quercetin-induced cell growth inhibition and apoptosis in human lung carcinoma cells. *J. Biol. Chem.*, 279, 53, 55875–55885, 2004.
12. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T. *et al.*, SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, 42, W252-8, 2014.
13. Delano, W.L., The PyMOL molecular graphics system, 2001, http://www.pymol.org.
14. Gasteiger, E., Hoogland, C., Gattiker, A. *et al.*, Protein identification and analysis tools on the ExPASy server, in: *The proteomics protocols handbook*, J.M. Walker (Ed.), Humana Press, Totowa, 2015.
15. Prabi, L.G., Color protein sequence analysis, 1998, https://npsaprabi. ibcp.fr/cgi- bin/npsa_automat.pl?page=/NPSA/npsa_color.html.
16. Costantini, S., Colonna, G., Facchiano, A.M., ESBRI: a web server for evaluating salt bridges in proteins. *Bioinformation*, 3, 137–138, 2008.
17. Roy, S., Maheshwari, N., Chauhan, R. *et al.*, Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. *Bioinformation*, 6, 8, 315–319, 2011.
18. Geourjon, C. and Deleage, G., SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.*, 11, 681–684, 1995.

19. Laskowski, R.A., MacArthur, M.W., Moss, D.S. *et al.*, PROCHECK: a program to check the stereo chemical quality of protein structures. *J. Appl. Cryst.*, 26, 283–291, 1993.
20. Wallner, B. and Elofsson, A., Can correct protein models be identified? *Protein Sci.*, 12, 1073–1086, 2003.
21. Colovos, C. and Yeates, T.O., Verification of protein structures: patterns of non-bonded atomic interactions. *Protein Sci.*, 2, 1511–1519, 1993.
22. Eisenberg, D., Luthy, R., Bowie, J.U., VERIFY3D: assessment of protein models with three- dimensional profiles. *Methods Enzymol.*, 77, 396–404, 1977.
23. Jayaram, B., Active site prediction server, 2004, http://www.scfbio-iitd.res.in/dock/ActiveSite.jsp.
24. National Center for Biotechnology Information, Pubchem., 2017, https://pubchem.ncbi.nlm.nih.gov/.
25. Hui, S.L. and Yang, Z., BSP-SLIM: A blind low-resolution ligand-protein docking approach using theoretically predicted protein structures. *Proteins*, 80, 93–110, 2012.
26. Kumar, S., Tsai, C.J., Ma, B. *et al.*, Contribution of salt bridges toward protein thermo-stability. *J. Biomol. Struct. Dyn.*, 1, 79–86, 2000.
27. Kumar, S. and Nussinov, R., Salt bridge stability in monomeric proteins. *J. Mol. Biol.*, 293, 1241–1255, 2009.
28. Kumar, S. and Nussinov, R., Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys. J.*, 83, 1595–1612, 2002.
29. Parvizpour, S., Shamsir, M.S., Razmara, J. *et al.*, Structural and functional analysis of a novel psychrophilic b-mannanase from *Glaciozyma Antarctica* PI12. *J. Comput. Aided Mol. Des.*, 28, 6, 685−698, 2014.
30. Murugesan, D., Ponnusamy, R.D., Gopalan, D.K., Molecular docking study of active phytocompounds from the methanolic leaf extract of *vitexnegundo*-against cyclooxygenase-2. *Bangladesh J. Pharmacol.*, 9, 2, 146–53, 2014.
31. Balamurugan, V. and Balakrishnan, V., Molecular docking studies of *Moringaconcanensis*nimmo leaf phytocompounds for brain cancer. *Res. Rev.: J. Life Sci.*, 8, 1, 26–34, 2018.
32. Santhanakrishnan, D., Sipriya, N., Chandrasekaran, B., Studies on the phytochemistry, spectroscopic characterization and antibacterial efficacy of Salicornia Brachiata. *Int. J. Pharm. Pharm. Sci.*, 6, 6430–6432, 2014.
33. Kasilingam, T. and Elengoe, A., *In silico* molecular modeling and docking of apigenin against the lung cancer cell proteins. *Asian J. Pharm. Clin. Res.*, 11, 9, 246–252, 2018.
34. Ashwini, S., Varkey, S.P., Shantaram, M., *Insilico* docking of polyphenolic compounds against Caspase 3-HeLa cell line protein. *Int. J. Drug Dev. Res.*, 9, 28–32, 2017.
35. Chakrabarty, N., Hossain, A., Barua, J., Kar, H., Akther, S., Al Mahabub, A., Majumder, M., *Insilico* molecular docking of some isolated selected compounds of lantana camera against Alzheimer's disease. *Biomed. J. Sci. Tech. Res.*, 12, 2, 9168–9171, 2018.

36. Supramaniam, G. and Elengoe, A., *In silico* molecular docking of glycyrrhizin and breast cancer cell line proteins, in: *Plant-Derived Bioactives*, 1, 575–589, 2020.
37. Elengoe, A. and Sebestian, E., *In silico* molecular modelling and docking of allicin, epigallocatechin-3-gallate and gingerol against colon cancer cell proteins. *Asia Pac. J. Mol. Biol. Biotechnol.*, 4, 2851–67, 2020.