DRUG DESIGN BY HOMOLOGY MODELING TO INHIBIT BREAST CANCER TARGETING PIGF PROTEIN

Abstract

The imbalance between apoptosis and cell proliferation or over-expression of a particular protein and deviation in biological pathways leads to cancer. Several types of cancers exist in which breast cancer is known since ancient times. 5-10% of breast cancers are due to inherited genetic disposition. Obesity, lack of exercise, alcoholism and exposure to ionizing radiations are a few risk factors for the development of breast cancer. In recent research, January 2022, in the United States, around 3.8 million women were identified with breast cancer history. PIGF (Placental growth factor), an angiogenic protein which is a subfamily of VEGF (Vascular endothelial growth factor)has a vital role in breast cancer. We identified P49763 protein having 221 amino acids as the target protein. For the FASTA sequence, the UniProtKB server was used. Search engine tools that were used to search the template include NCBI, Phyre2, JPred, and HH Pred. PDB Id 1FZV was identified as the suitable template with 92.06% and low E- Value with 65% of query coverage. The clustalW was used for the alignment of the target and template sequence. The Aligned sequence was submitted to the SWISS-MODEL server for the 3D model. The model was verified by ERRAT, PROCHECK, and VERIFY 3D. The Favorable region in the Ramachandran Plot was found to be 91.6%. The ERRAT portal showed overall quality factor of 96.97%. The verify 3D showed less than 80% of amino acids that have scored ≥ 0.1 in the 3D/1D profile. The 3D model built was docked with VEGFR 1(a natural substrate) and identified that CYS34, ASP71, GLY70, SE56, GLU27, PHE25, TRP29, and LEU221 amino acids play a key role in

Authors

Navaneetha Nambigari
Assistant Professor
Dept of Chemistry
University College of Science
Osmania University, Hyderabad, India
nitha379@gmail.com

Pushpanjali Pendyala
Lecturer
Department of Biotechnology
TSWR Degree and PG college
Mahendra hills, Hyderabad, India.
pushpamarvadi@gmail.com

Glory Prathiba
TSWR Degree and PG college-Mahendra hills
Hyderabad, India.
kopperaglory325@gmail.com

pathway progress. Inhibition of interaction of any of these amino acid residues may hinder cancer development.

Keywords: Breast cancer, Angiogenesis, PIGF, VEGF, VEGFR1, Homology Modeling

I. INTRODUCTION

Breast cancer has highest prevalence in women.The most diagnosed disease in US is breast cancer after lung cancer. The latest statistic update provided by American Society about breast cancer for women estimated to be 297, 790including the number of new cases and death [1-2]. Over the last 26 years, incidents rate of Breast cancer in females is increased by 39.1% [3]. Breast cancer originates in different tissues of the breast. If cancer starts in Lobules, the glands that make breast milk, it is known as lobular cancer. If it is caused in ducts that connect lobules and nipples, it is known as Ductal cancer. Cancer that starts in the nipple is calledPaget disease and breast cancer which originates in stroma it is known as the Phyllodes tumor. If cancer starts in blood vessels and lymph vessels of breast, then it is known as Angiosarcoma [4]. Risk factors of the breast cancer include ageing, Mutations (BRCA1, BRCA2), Reproductive history, Having Dense breast, Family history, Exposure to Radiation therapy or Drugs, Physical Inactiveness, Alcohol consumption, Obesity, Hormonal therapy etc [5,6,7]. Symptoms include lumps in breast or underarm, Swelling of part of the breast, Irritation of breast skin, Discharge of blood from nipples, Change in the shape or size of the breast and Pain in any area in the breast etc.

Signal transduction at cellular level is a fundamental process in development and progression of cancer. Changes in various cell signaling pathways promote tumor growth. Mutations associated with hereditary breast cancer include BRCA1, BRAC2 and TP53 genes [8].Breast cancer is classified into different sub types based on involvement of estrogen receptor(ER), Progesterone receptor(PR), and HER2 receptor. Activity of HER2 receptors effectssignaling of other pathways such as MAPKs, GSK-3, PI3K/Akt/mTOR pathways accelerating the progression of breast cancer. PI3K/Akt/mTOR signaling pathways involved in growth, proliferation, survival and metabolism are associated with several diseases and syndromes including breast cancer[9]. Most of the clinically available drugs target any of the key factor in the pathway and thereby effect cancer progression. Evista (raloxifene hydrochloride), Soltamox (tamoxifen citrate), Elacestrant, Abemaciclib are few examples of such drugs used to treat breast cancer whose mode of action is through binding to their receptors and inhibit the signal transduction functioning in cancer.

Targeting angiogenic factors such as vascular endothelial growth factor (VEGF) or transcription factors are found to be effective treatment options for several types of cancers. Hormones and their receptors also play an important role in the pathogenesis of breast cancer. Among several factors, the family of vascular endothelial growth factors, are the most potent pro-angiogenic factors.They comprise secreted proteins including VEGF-A, VEGF-B, VEGF-C, VEGF-D and placental growth factor (PlGF)[10] . Placental growth factor (PIGF) selectively binds to the tyrosine kinase receptors which are present on the endothelial cells. The natural substrate of PIGF is VEGFR1 which on phosphorylation activates and modulates the Akt, PI3K pathways. The activation of these signal transduction pathwayspromotes survival of endothelial cells and enhancement of vascular permeability thereby promoting angiogenesis around tissues. Blocking of these pathways will hinder blood supply and essential nutrients for the growth of cancer cells leading to arrest of cancer progression.

II. METHODOLOGY

Using Universal Protein Knowledgebase(Uniprot) database, we identified PIGF proteins the Target sequence with 221 amino acids and with no is forms and Uniprot ID is P49763.Homology modeling was used to construct the 3D Structure of a protein. Pair wise Alignment was done by CLUSTAL W and 3D structure was generated by BIOVIA discovery studio 3.0. SWISS MODEL Secondary Structure analysis by homology modeling, PROCHECK was used to validate the model, ERRAT from the structural Analysis and verification Server (SAVES) was utilized to analyses structure. Ramachandran plots were used to understand the distribution of the (φ, ψ) torsion angles of the protein backbone. Pro SA plot was used to study amino acid residues. NCBI blast for active site identification.

III. RESULTS AND DISCUSSION

1. **Homology Modeling:** Homology Modeling is also known as the Comparative Modeling of protein. For template sequence, the target protein was retrieved from protein BLAST in NCBI, and the PDB ID was identified as 1FZV. This ID was considered as a Template sequence as it gave accurate lowest e-value scores with sixty percentage. The lowest e-values of proteins are considered as the template. The template selection also works on particular criteria such as Identity, sequence, similarities, and e- values. The target sequence was then submitted to search engine tools such as BLAST for sequence identification, HH Pred for remote protein homology detection and J Pred for secondary structure identification. All these criteria were analyzed for identification of the template for protein structure prediction as shown in Table 1.

Table 1: Template Search

S. No	SERVER	ID	E-value
1	NCBI BLAST	1 FZV	1.00 e-79
2	J PRED	1 FZV	6.00 e-64
3	HH PRED	1 FZV	3.60 e-18

The 1FZV template protein is a chain A, placenta growth factor and it has a 65% query Coverage, and with 92.06%.

2. **Alignment:** Target sequence (P49763) and Template sequence (1FZV) were submitted for the Pair-wise alignment by CLUSTAL W [11]. The Alignment identified the regions of similarities of target and template sequences as shown in the Figure.1. The blue color letters represent the similarities of target and template sequences. There are 221 amino acids in which 131 amino acids are found to be similar, making 60% of strong similarities between target and template sequences. The similarity region starts 69^{th} Amino acid and extends till 81^{th} amino acid. Another region spans from 199^{th}amino acid to 221^{th} amino acid. Further, these alignments were submitted to BIOVIA discovery studio 3.0 [12] to generate 3D stucture as shown in Figure 2. The Biovia Discovery studio predicts the secondary structure of a protein and is used for stimulating small molecules and macromolecule systems it identifies the three-dimensional fold of a protein, it understands

the spatial orientation and interactions of active site residues. SWISS MODEL [13 and 14] is an web based server for automated comparative modeling of 3D protein structures, it helps in building proteins homology models at different levels of complexity. 3D model was then assessed for it's quality by Q-mean scoring Function. Q mean -Qualitative model energy Analysis is represented in Figure 3. SWISS-MODEL relies on Q-mean scoring function and it uses statistical potentials of mean force to make global and per residue quality estimates. The local quality estimates are enhanced by pair wise distance constraints that represent ensemble information from all template structures found, Q mean Z-scores around zero indicate good agreement between the model and experimental structures of similar size. Q mean Z score was found to be 0.76 which is in line with the prediction.

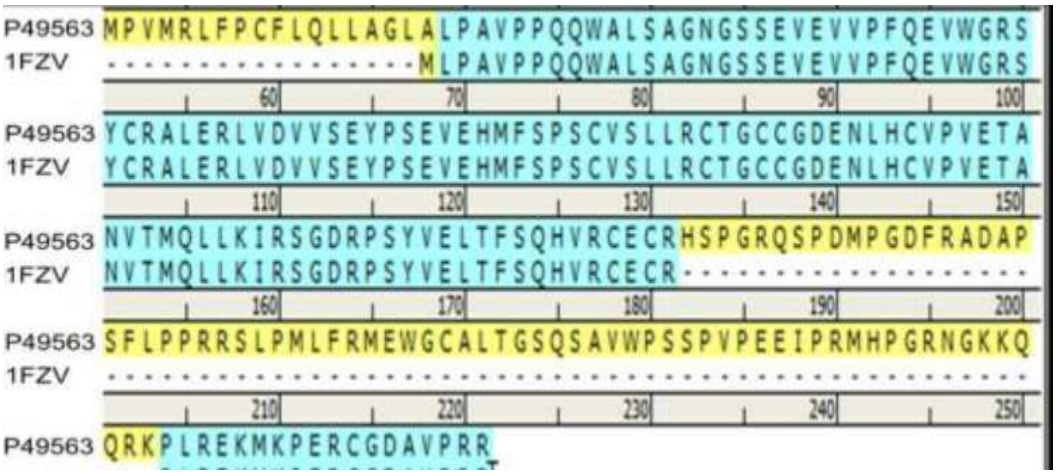

Figure 1: Alignment of P49763 (Target protein) and 1FZV (Template Protein)

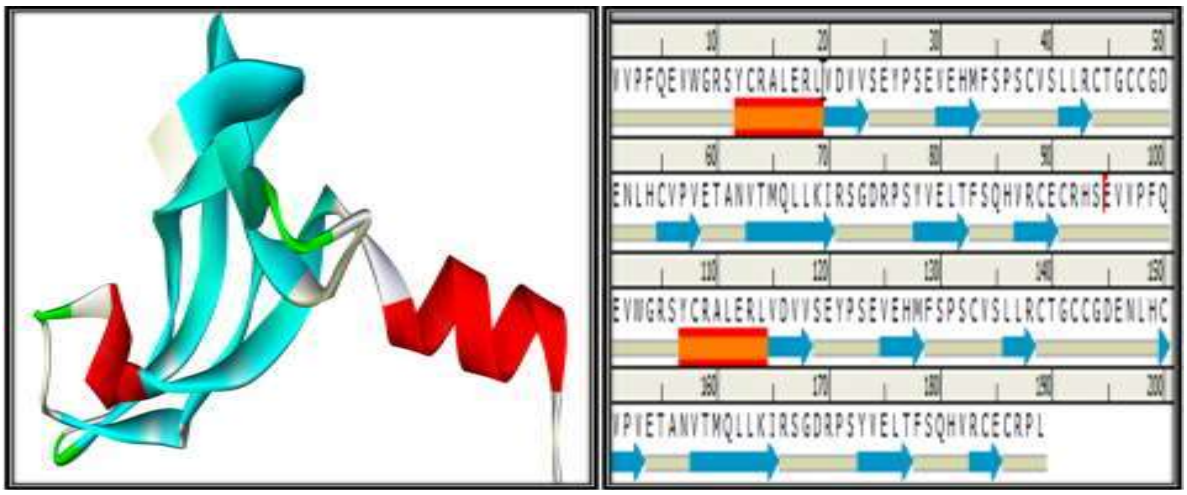

Figure 2: 3D model of protein secondary structure as predicted by Biovia discovery studio 3.0

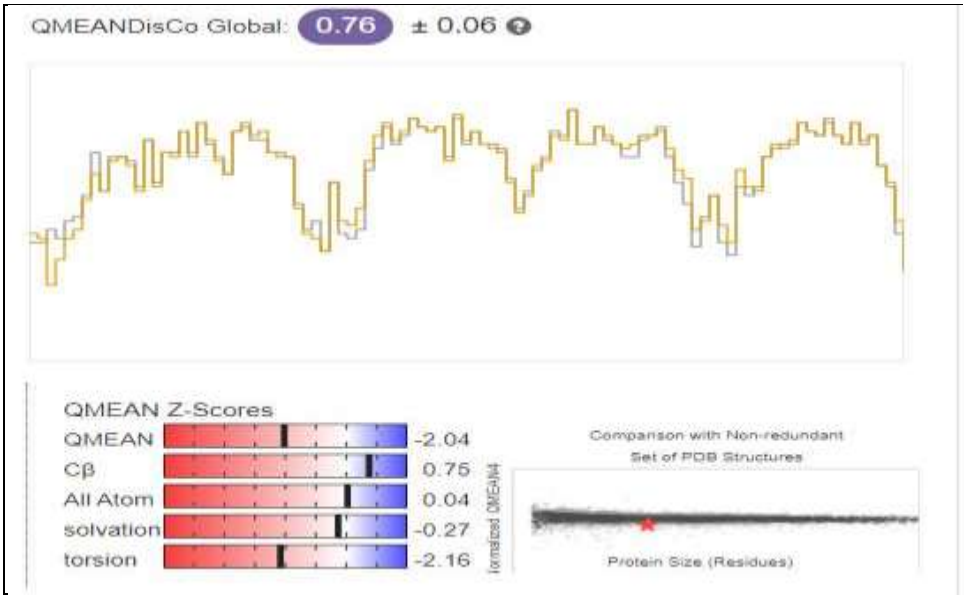

Figure 3: Q-mean score of 3D structure of a proteinP49763

3. **Secondary Structure Analysis:** The 3D structure was generated by the homology modeling, the structure was found to comprise of Five beta sheets, one gamma sheet, one beta hairpin, Two helices, five strands and three disulphide bonds along with amino acid chain lengths as shown in Figure 4.

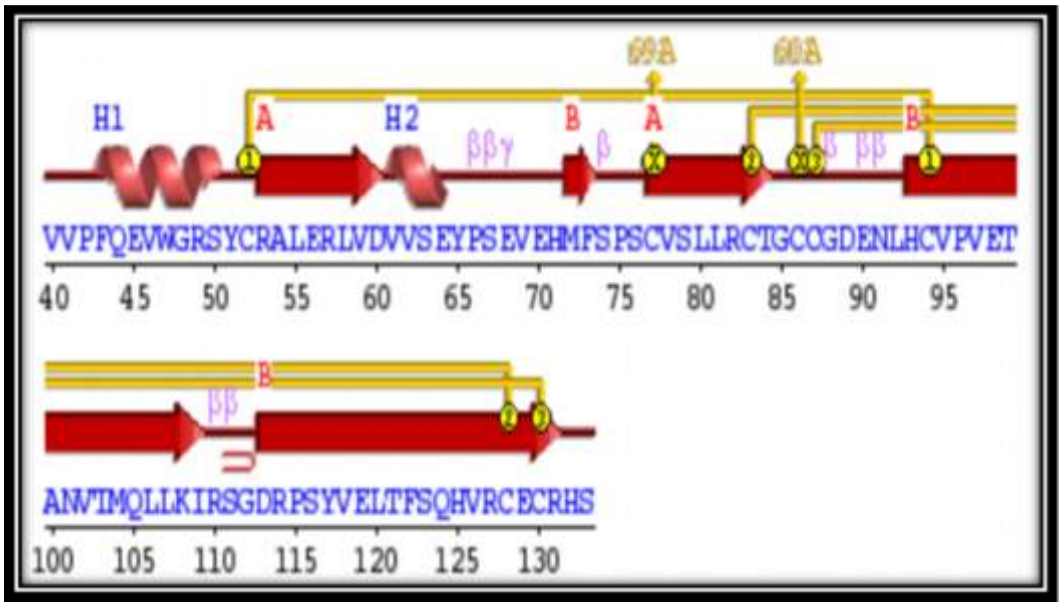

Figure 4: Secondary Structure of target protein-P49763

4. **Validation:** PROCHECK was used to validate the model, ERRAT was used from the structural Analysis and verification server (SAVES). Ramachandran plots (Figure 5) report the distribution of the (φ, ψ) torsion angles of the protein backbone. This plot explains the number of residues belonging to "Outlier", "allowed" and "favored' regions.

The red color in plots represents the most favorable region and the yellow points on additionally allowed as shown in Figure.5. It was observed that 91.6% are in the favored regions, and 7.8% are in additional allowed regions. The total number of residues is 189 of which 166 are found to be non-glycine and non-proline residues. The plot statistics are given in Table.2. The observation indicates that the protein was stereo chemically appropriate and was generated after energy minimization.[15, 16, 17]. Further, the Pro SA plot [18] is used to analyses local model quality, represented in Figure 6. On analyzing, the Pro SA plot shows that most of the amino acid residues have negative energy indicating that the PIGF protein has good local quality in the structure. This Pro SA plot z-score value is -4.68 that falls in the NMR regions. ERRAT is used to check an overall quality factor for non-bonded atomic interactions in which higher scores with minimum error confirms higher quality. The overall quality is 96.970 as shown in Figure 7.

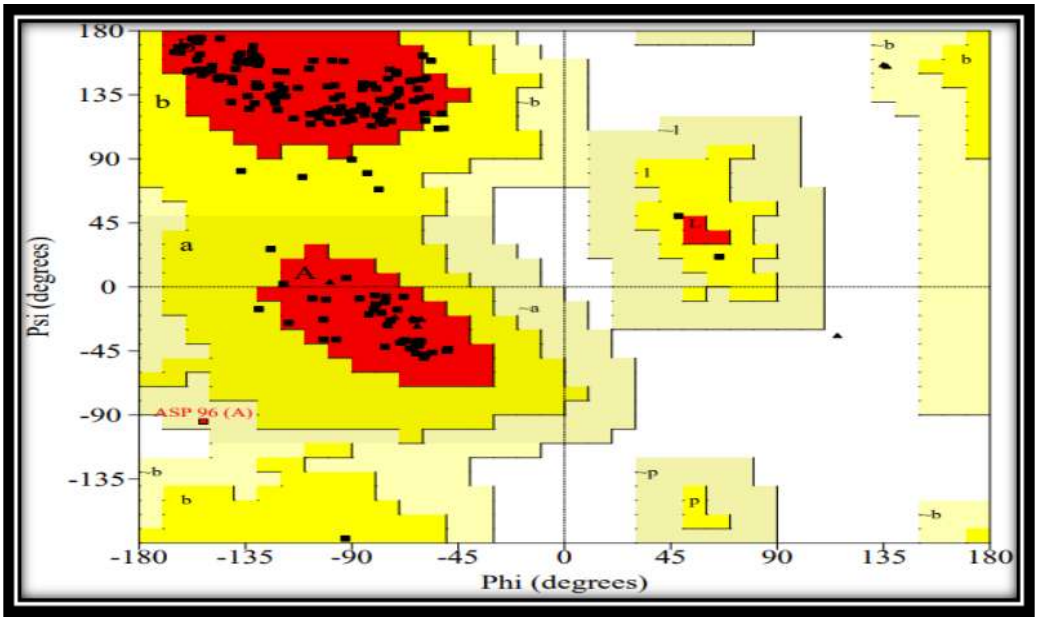

Figure 5: Ramachandran plot of the 3D protein.

Table 2: Ramachandran plot of statistics

Plot statistics

Residues in most favoured regions [A,B,L]	152	91.6%
Residues in additional allowed regions [a,b,l,p]	13	7.8%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.6%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	166	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	8	
Number of proline residues	11	
Total number of residues	189	

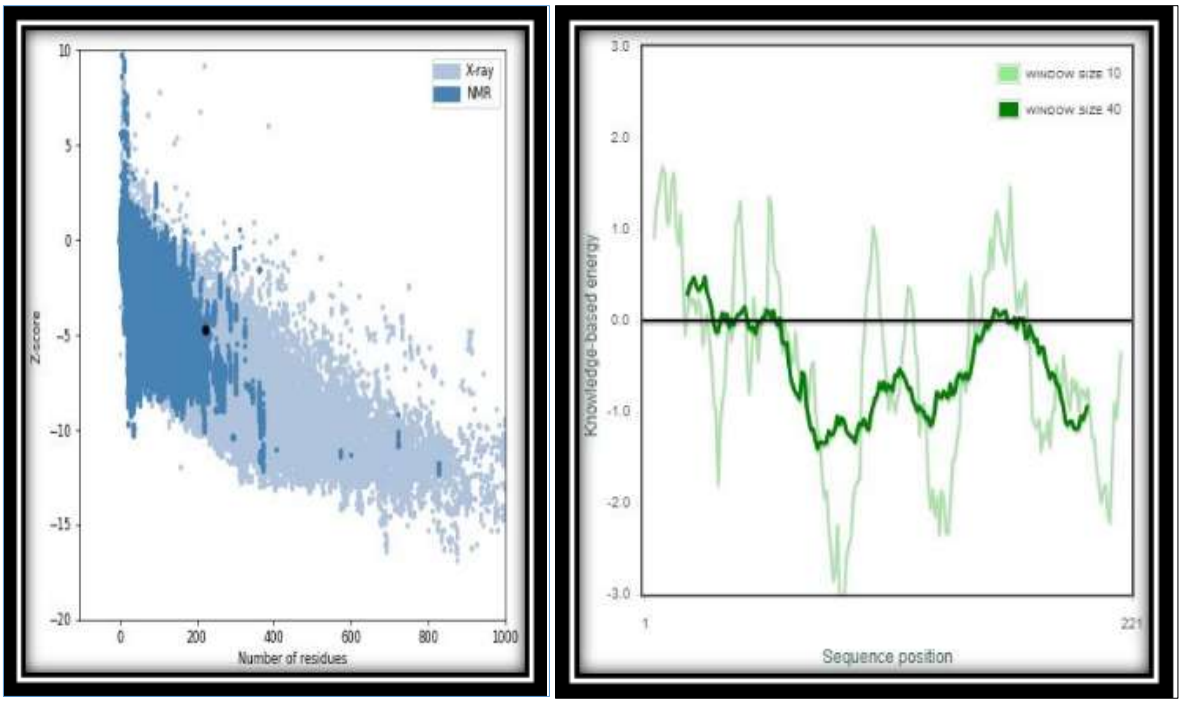

Figure 6: A) ProSAplot. Black spot represents the 3D model falls in the NMR regionwith the Z- score= -4.68. B) ProSA plot Energy Profile.

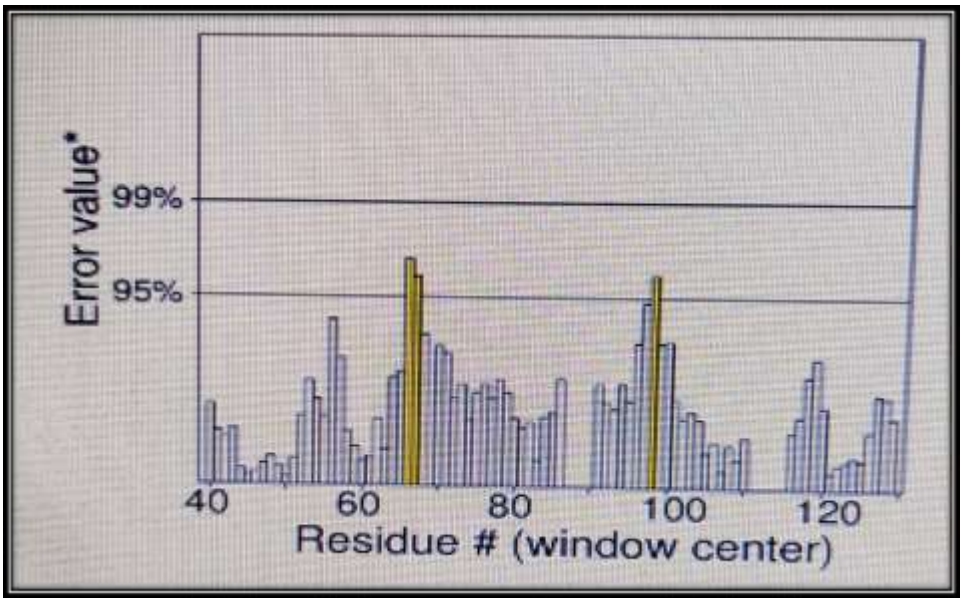

Figure 7: ERRAT Plot with overall quality Factor**: 96.970

5. **Active Site Identification:** The conserved domains of PIGF were also characterized by NCBI blast as shown in Figure 8. Further, the binding site domains of the PIGF protein were also identified. The protein–protein docking sites are CYS34, ASP71, GLY70, SE56, GLU27, PHE25, TRP29, LEU221 are found to be conserved. The amino acids residues are shown as ball and stick as shown in Figure 9 [19]

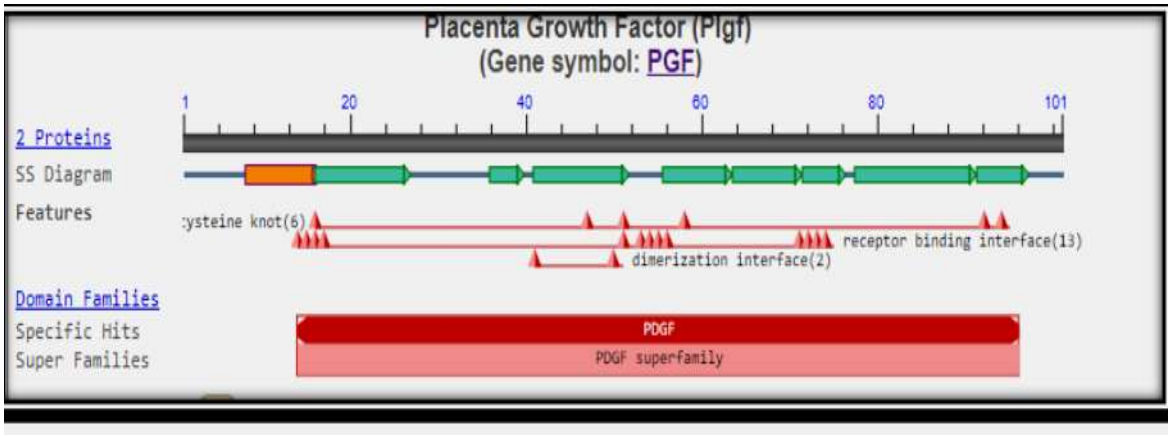

Figure 8: Conserved domain of PIGF Protein.

Figure 9: Interaction showing Protein Protein Docking

IV. CONCLUSION

Earlier studies showed that PIGF Protein expression is higher in breast cancer and were found to be associated with survival without recurrence of the disease. Hence, we identified PIG Fas target protein to build 3D Model through Homology modeling, Further, we searched for Template sequence using search engine tools such as BLAST, H pred, J pred and identified PDB ID:1FZV as the template protein. The template and target protein were assessed through several servers such as CLUSTAL W, SWISS MODEL. BIOVIA DISCOVERY STUDIO, PROCHECK, Pro SA (A,B) and ERRAT to understand and assess the template protein. The results are promising proving that this the potential protein for development of drug that interacts and hinders the signaling pathway and thereby cancer. Rama Chandran plot scores revealed that the model is reliable. Further docking studies concluded that CYS34, ASP71,GLY70, SE56,GLU27,PHE25,TRP29, LEU221 amino acids of PIGF are involved in interactions that are required for PIGF protein to bind to it's receptor. The binding signals downstream effectors resulting in inhibition of target genes. Further, understanding the receptor domain docking site of PIGF helps in identifying new molecular entities as PIGF antagonist.

REFERENCES

[1] American Cancer Society. Breast Cancer Facts and Figures 2022-2024. Atlanta: American Cancer Society, Inc. 2022.
[2] C. DeSantis, J. Ma, L. Bryan and Jemal A. "Breast cancer statistics", CA Cancer J, vol 64, 2014, pp52–62.
[3] American Cancer Society. Breast Cancer Facts and Figures, 2015- 2016. Atlanta: American Cancer Society, Inc. 2015.
[4] NL. Henry, PD. Shah, I. Haider, PE. Freer, R. Jagsi and MS Sabel. Chapter 88: Cancer of the Breast. In: JE. Niederhuber, JO. Armitage, JH. Doroshow, MB. Kastan and JE Tepper, eds. Abeloff's Clinical Oncology. 6th ed. Philadelphia, Pa: Elsevier; 2020.
[5] RT. Chlebowski. Factors that modify breast cancer risk in women. UpToDate. 2021. Accessed at https://www.uptodate.com/contents/factors-that-modify-breast-cancer-risk-in-women on August 25, 2021.
[6] J. Kruk and H. Aboul-Enein. Environmental exposure, and other behavioral risk factors in breast cancer. Current Cancer Therapy. Reviews.vol 2(1) 2006, pp 3–21.
[7] A. Howell, AS. Anderson, RB. Clarke, SW. Duffy, DG. Evans, M. Garcia-Closas, AJ. Gescher, TJ. Key, JM. Saxton and MN. Harvie. Risk determination and prevention of breast cancer. Breast Cancer Res. Vol 16(5, 2014, :446.
[8] L. Melchor andJ. Benítez. The complex genetic landscape of familial breast cancer. Human Genetics. 132(8)2013, pp845–863.
[9] T. Mukohara T. PI3K mutations in breast cancer: prognostic and therapeutic implications. Breast Cancer (Dove Med Press). Vol 7, 2015, pp111–123.
[10] JM. Castañeda-Gill and JK Vishwanatha. Antiangiogenic mechanisms and factors in breast cancer treatment.J. Carcinogenesis. Vol 15.1, 2016.
[11] JD. Thompson, DG. Higgins and TJ. Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. Vol 22(22) 1994, pp4673 4680.
[12] A. Sali, I. Potterton, F. Yuan, H. Vlijmen andM. Karplus. Evaluation of comparative protein modelling by MODELLER.Proteins, Vol 23, 1995, pp318 -326.
[13] N. Guex and MC Peitsch. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. Vol 18(15), 1997, pp2714-2723.
[14] RA. Laskowsky, MW. MacArthur MW, DS. Moss and JM Thornton. PROCHECK: a program to check the stereo chemical quality of protein structure.J Appl Crystallography, Vol 26, 1993, pp283- 291.
[15] AL. Morris, MW. MacArthur, EG. Hutchinson andJM. Thornton. Stereo chemical quality of protein structure coordinates. Proteins, Vol 12, 1992, pp345-364.
[16] GN. Ramachandran, C. Ramakrishnan and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. J MolBiol, Vol 7, 1963, pp 95-99.
[17] Oliviero Carugo and Kristina Djinovic-Carugo. Half a century of Ramachandran plots, Acta Crystallographic Section D, Vol 69(8), 2013, pp1333-1341.
[18] M. Wiederstein and MJ. Sippl. ProSA- web: interactive web service for the recognition of errors in the three –dimensional structure of proteins. Nucleic Acids res, Vol 35, 2007),pp407- 410.
[19] M. Schapira, M. Totrov and R. Abagyan. Prediction of binding energy for small molecules, peptides and proteins. J. Mol. Recognit. Vol 12(3), 1999, 177-190.

