Readers Insight
Publisher of open access books and journals
PJMLS
Online ISSN: 2707-4471. Print ISSN: 2707-448X

Research Article	Pak-Euro Journal of Medical and Life Sciences
DOI: 10.31580/pjmls.v3i4.1730	Copyright © All rights are reserved by Corresponding Author
Vol. 3 No. 4, 2020: pp. 187-194	
www.readersinsight.net/pjmls	
Submission: November 19, 2020	**Revised:** December 18, 2020 **Accepted:** December 30, 2020

ANALYSIS OF SEROLOGICAL PROTEINS OF BREAST CANCER PATIENTS BY USING BIOINFORMATICS TOOLS

Sabeera Afzal[1], Zunera Tanveer[2]*, Sumaira Awan[1], Rimsha Rasheed Malik[1], Madiha Sana[3], Khawar Iqbal[4], Syed Haseeb Shah[1], Mariam Salim[5], Mariam Khaliq[1]

[1]**Center for Advanced Studies in Vaccinology & Biotechnology (CASVAB)**, University of Balochistan, Quetta, Pakistan.

[2]**Department of Physiology**, Bolan University of Medical and Health Sciences, Quetta, Pakistan

[3]**Department of Mathematics**, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

[4]**Department of General Surgery,** Postgraduate Medical Institute Quetta, Pakistan

[5]**Rahat Medical Complex,** Quetta, Pakistan

***Corresponding Author:** Dr. Zunera Tanveer Email: zunerambbs@gmaill.com

Abstract

The primary and most important aim of using bioinformatics tools in this study is to better understand and predict about the proteins and their 3D structure, physicochemical properties and the amino acids sequences. These information about a desired protein is important in the field of science and bioinformatics which made it possible for the scientists to learned about the proteins in detail and compare to experimental tools it is more accurate and explains proteins more accurately. In the world many scientists are working on these tools so they can predict different structures and functions of proteins. Breast cancer (BC) is one of the most common diseases globally affecting approximately 2.1 million women every year. Proteomics and Bioinformatics play a vital role in the better understanding of BC proteins with the help of different present bio tools. There were 70 total serum samples of positive BC patients and for the separation of proteins on the basis of their molecular weight. Bioinformatics was applied to all the detected protein's and their sequence of amino acids. The physiochemical properties show that the detected experimental and theoretical molecular weights of Mucin-1 (MUC-1), 60 kDa heat shock protein (HSP-60), G2/mitotic-specific cyclin-B1 (Cyclin B1), Receptor tyrosine-protein kinase erbB-2 (HER-2/neu) and Myc proto-oncogene protein (C-Myc) was 115 & 119, 59 & 60, 50 & 48,52 & 48.8 and 130 & 135 KDa respectively. Putative transmembrane domains were found highly in HSP60 and Cyclin B1, less and moderately in MUC-1 and HER-2/neu and not shown in C-Myc in end result. The similarity between physio chemical properties of two or more protein can help the biologists to further work on those proteins specifically for BC diagnostics.

Keywords: *Breast cancer, Bioinformatics tools, SDS-PAGE, Serological Proteins*

INTRODUCTION

In molecular biology, biochemistry and other fields of biology, an emerging and interdisciplinary subject known as bioinformatics or computational biology plays an important role and the researcher or scientist can study better about the living things by using multiple software's, computer technology and algorithms. These software's help to better understand the multifaceted biological or genetic system such as tissues, human body system and cells (1). There are several tools are available for example, a database

known as Protein Data Bank (PDB) which is developed at 1971 and is used for the three-dimensional (3D) structural data of large biological molecules (such as protein and nucleic acid) and have all the information about different proteins (2).

Breast cancer (BC) is one of the most common diseases globally affecting approximately 2.1 million women every year. BC does occur in men but it is very rare accounts for 1% from all the BC patients (3). According to WHO in 2018 cancer was the second leading cause of death worldwide and in other types of cancers, BC was the second most diagnosed cancer worldwide after lung cancer (4). Breast cancer results from mutation in essential genes and these mutations occurs because of environmental and genetic factors. From past some years the risk of having BC in a lifetime is increased in Pakistan where every 1 out of 9 woman has a lifetime risk of being diagnosed with BC. Because of late diagnosis and less facilities, the mortality rate of BC is higher in Pakistan compare to other countries. Also, the worst problem in Pakistan is that there is not any available data of accurate occurrence rate, mortality rate and new diagnosis cases rates per year, the only details which are present is the data collected from the hospital (5, 6). Studies shows that Autoantibodies (AABs) are related with Tumor Associated Antigens (TAAs) and to diagnose BC, AABs identification is used as novel method and it can predict the tumor when the size is so small to detect by other methods. In BC, tumor antigen CA 15-3 is a serum soluble glycoprotein and it is the most commonly used marker but it is limited because of its low sensitivity and specificity (7). Some other protein which has been observed to be associated with BC includes p53, c-myc, mucin 1(MUC-1), cyclin B1, cyclin D1, heat shock proteins (HSP-27, HSP-60, and HSP-90), human epidermal growth factor receptor 2 (HER2) and cancer-testis antigens (NY-ESO-1) (8).

Proteomics and bioinformatics play a vital role in the better understanding of BC proteins with the help of different present tools. There are multiple databases and analytical tools provided by a multidisciplinary team at the swiss institute of bioinformatics. It can be access through the world wide web server known as ExPASy (the Expert Protein Analysis System). In the world of life science ExPASy was the first WWW server, started to function in 1993. In ExPASy there are numerous tools which include UniProt, ProtParam, ProtScale, SwissProt and TrEMBL etc. which helps in the prediction of the isoelectric point of proteins, physiochemical properties, their Primary, secondary and tertiary structure and their functions, transmembrane domain and most importantly its molecular weight (9, 10).

UniProt (the Universal Protein Resource) delivers all the functional data, information and protein sequence and accession number regarding the specific protein with a high quality and easily accessible resource. That information about the specific protein can be further processed in ProtParam which is an important tool of ExPASy which helps in the better understanding of protein sequence and calculate its Physio-chemical properties for instance the atomic composition, theoretical pI, molecular weight of protein, aliphatic index, instability index, the extinction coefficient, the estimated half-life and composition of amino acids etc (11, 12).

ProtScale is another important tool which have some 50 predefined scales for protein like Doolittle and Kyte hydrophobicity scale. These amino acids scales are defined by a specifically given numerical value assigned to amino acids based on their structure and functions. ProtScale helps to calculate or compute and show the profile by the given amino acid scales on a specific protein.

Other than ExPASy another excellent and important tool is PSIPRED used for the secondary structure prediction of specific protein. National Center for Biotechnology Information (NCBI) and Protein Data Bank (PDB) was used to predict the 3D model of the Specific protein (10-12).

METHODOLOGY

SAMPLE COLLECTION

There were 70 total serum samples of positive BC patients taken from Bolan Medical Complex, Rahat hospital and LINAR hospital loralai and a pool of 10 sera from healthy individual were taken as control samples. For the collection of samples falcon tubes without EDTA or anticoagulant were used. All the blood samples were centrifuged for 10 minutes at 13000 rpm for the separation of serum from blood and stored at -80°C in an Eppendorf tubes.

SDS-PAGE

For the separation of protein on the basis of their molecular weight a technique was used known as SDS-PAGE (Sodium dodecyl sulfate polyacrylamide gel electrophoresis). All the samples were performed at SDS-PAGE (13). There were two different gels were used known as resolving or Migration gel in which proteins are separated according to their molecular weight in gel matrix and Stacking gel in which samples are pour be prior it transfers to migration gel. In SDS-Page the protein who are high in charge and lower in weight transfer more rapidly compare to low charge and high molecular weight (14).

COOMASSIE BLUE STAINING

Coomassie brilliant blue staining **(**CBB) first discovered in 1926. It was used for the visualization of protein bands. There are a number of staining methods but Coomassie remains the most widely used method globally because of its simplicity, visual inspection, easy to operate and low cost compare to other methods e.g., Silver staining and Transilluminator apparatus (Lab net) was used to observed the pictures of gels (15, 16).

BIOINFORMATICS TOOLS

To predict the physiochemical properties, protein's secondary structure, transmembrane domains and 3D structures; different bioinformatics tools were used. The physiochemical properties of the proteins were predicted using ExPASy (ProtParam), protein secondary structures were predicted by PSIPRED, transmembrane domains were predicted by ExPASy (ProtScale) and protein 3D structures were predicted using NCBI, PDB (17).

RESULTS AND DISCUSSION

Total 70 BC patient's serum samples and 10 healthy individual's serum samples were taken and protein profiling was performed on those patients using SDS-PAGE. Five proteins (MUC-1, HSP-60, CYCLIN B1, C-MYC, HER-2/neu) were differently expressed in sera of 70 BC patients with different percentages. In addition, for further details variety of bioinformatics tools were used to predict the protein's physiochemical properties, proteins secondary structure, transmembrane domains and 3D structures.

Bioinformatics was applied to all the detected proteins and their sequence of amino acids. The physiochemical properties show that the theoretical molecular weights of Mucin-1 (MUC-1), 60 kDa heat shock protein (HSP-60), G2/mitotic-specific cyclin-B1 (Cyclin B1), Receptor tyrosine-protein kinase erbB-2 (HER-2/neu) and Myc proto-oncogene protein (C-Myc) is 119, 57, 48, 135 and 48.8 kDa correspondingly, however compare it with the detected experimental molecular weight which was 115, 59, 50, 130 and 52KDa respectively as shown in table 1. The theoretical iso-electric points (PI) were not done practically but according to ExPASy the PI Points of MUC-1, HSP-60, Cyclin B1, HER-2/neu and C-Myc is 6.96, 5.24, 7.09, 5.58 and 5.33 correspondingly.

Table I: Physiochemical properties of detected proteins on the basis of their degree of hydrophobicity

Name of Proteins	Accession Number	No. of Amino Acids	Theoretical Molecular weight (kDa)	Experimental Molecular weight (kDa)	Ext. coefficient	Theoretical PI	Estimated half-life	Instability Index	Aliphatic Index	Grand average of hydropathicity
MUC-1	P15941	1232	119	115	30495	6.96	1.9 hours	50.00	45.61	-0.501
HSP-60	P10809	547	57	59	14565	5.24	4.4 hours	28.39	100.71	-0.081
HER-2/neu	P04626	1233	135	130	132775	5.58	7.2 hours	56.33	81.20	-0.273
Cyclin B1	P14635	433	48	50	30620	7.09	30 hours	50.59	90.09	-0.239
C-Myc	P01106	439	48.8	52	29505	5.33	30 hours	92.23	66.42	-0.772

Accession numbers of proteins were taken from UniProt for further process. All the physiochemical properties are shown in Table I. Some proteins have the similar physiochemical properties such as MUC-1, Cyclin B1 and C-Myc which have similar ext. coefficient and MUC-1 and HER-2/neu have around equal numbers of amino acids and have similar molecular weight (Table I).

PREDICTION OF PUTATIVE TRANSMEMBRANE DOMAIN IN DETECTED AND IDENTIFIED PROTEINS

Bioinformatics tools significantly ease the study of proteomics and also it facilitates the identification and detection of new and better biomarkers. As shown in the (Figure 1), bioinformatics computation was applied on the detected proteins and their amino acids sequence to obtain the Putative transmembrane (TM) domains by using the scale Hphob. / Kyte & Doolittle developed in 1982. On the algorithms the hydrophobicity profile of

proteins made known at y-axis and location of the amino acids shown at x-axis. The calculated values should be above 1.5 for finding the transmembrane domain. Transmembrane domain was more found in HSP60 and CYCLIN B1, less and moderate in MUC-1 and HER-2/neu and not shown in C-MYC in end result. The identification of transmembrane domain in the selected protein is an important step to know more about the cell and its biological functions. Basically, PTM is any membrane-spanning protein domain which comes from nonpolar amino acid residue and it also holds the single alpha helix of a transmembrane protein. The concentration of protein in two dimensions is extreme in the membrane of cell where the transmembrane domains exists, which can majorly influence the functions and clustering (18).

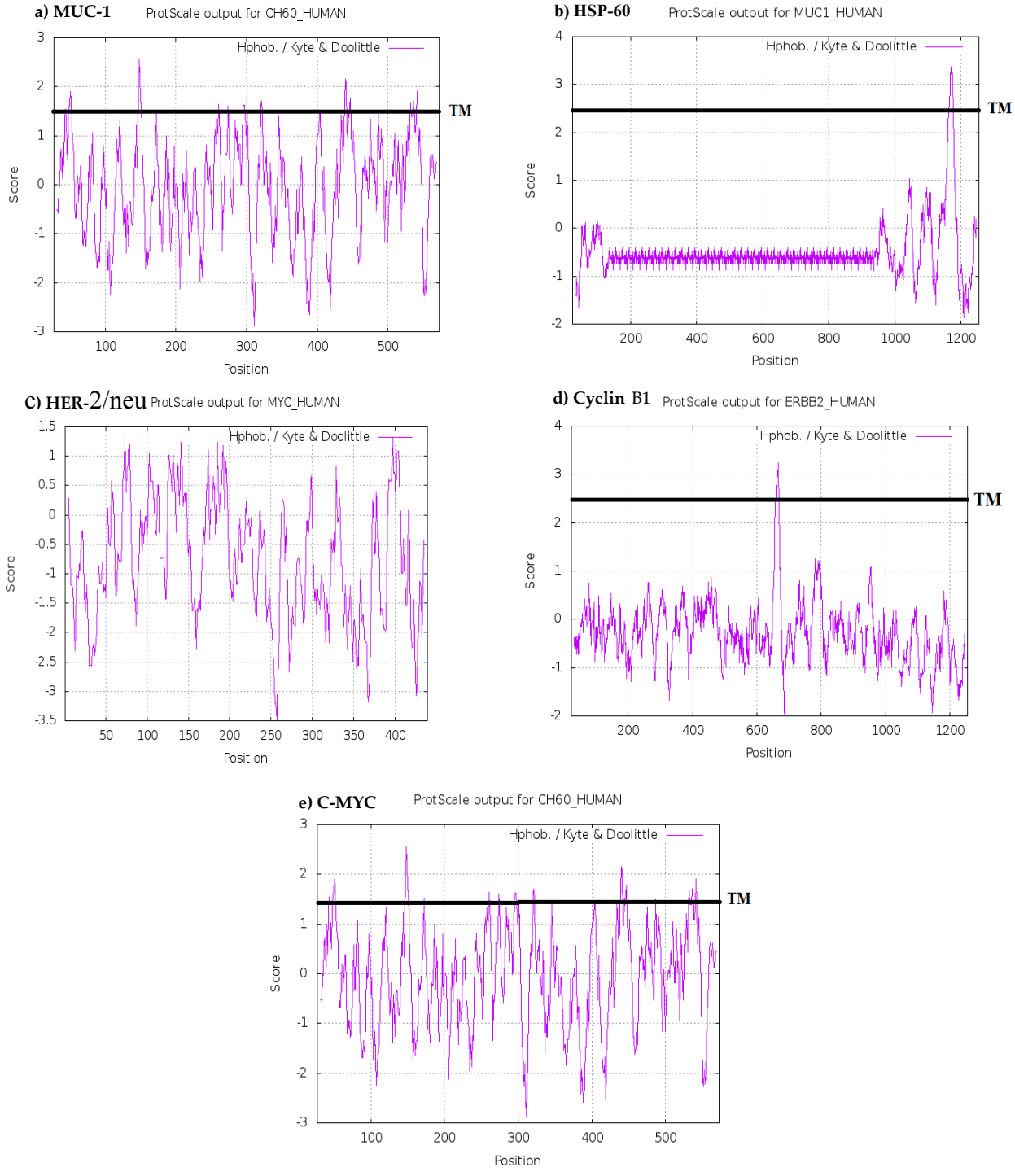

Fig. 1. Transmembrane Domains of a) MUC-1, b) HSP-60, c) Cyclin B1, d) HER-2/neu and e) C-MYC were obtained using the scale Hphob. / Kyte & Doolittle (1982).

Readers Insight
Publisher of open access books and journals
PJMLS
Online ISSN: 2707-4471. Print ISSN: 2707-448X

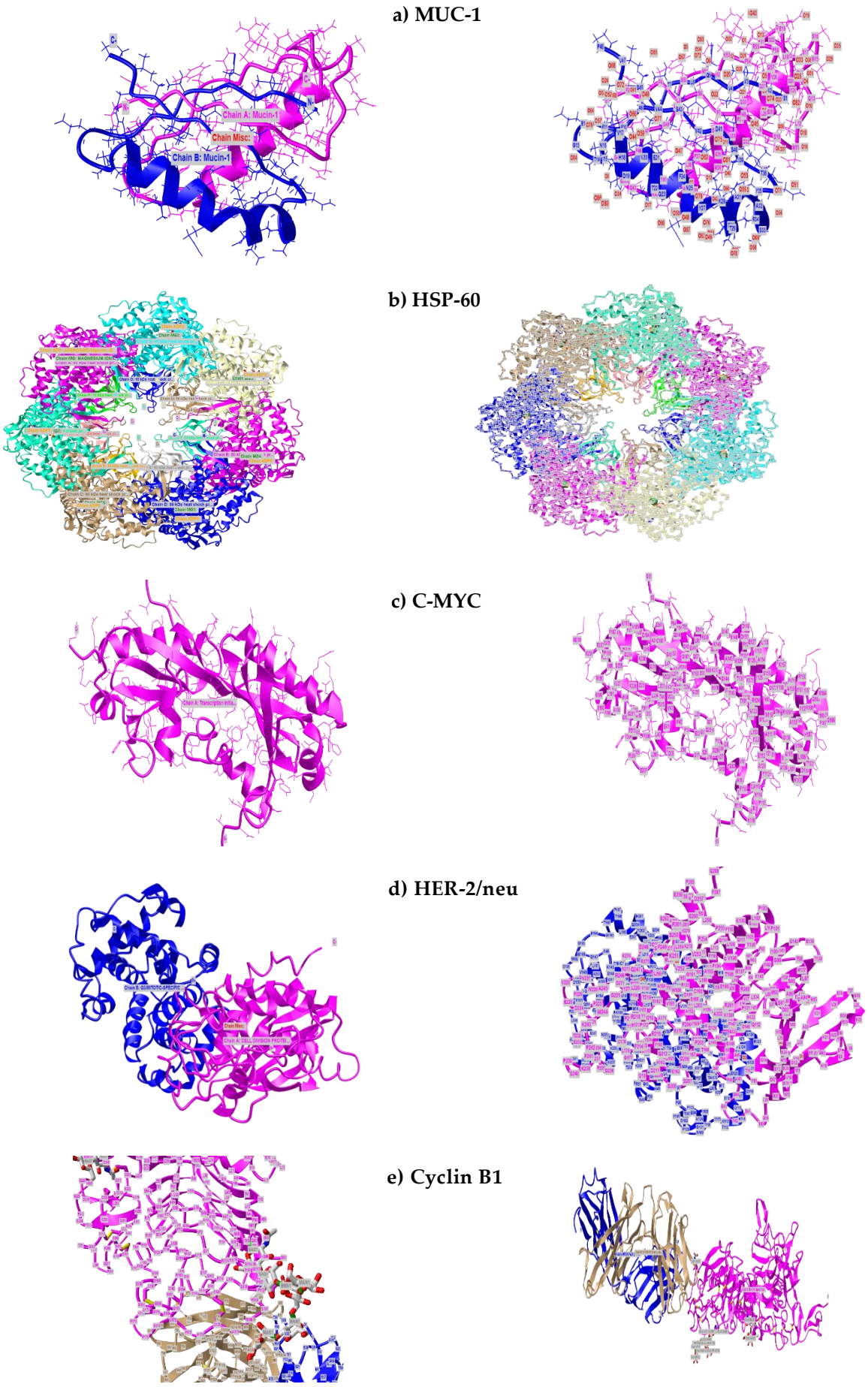

Fig. 2. 3D structures of MUC-1, HSP-60, Cyclin B1, HER-2/neu and C-Myc were predicted using NCBI and PDB

3D STRUCTURES OF DETECTED PROTEIN

NCBI, PDB was used to predict the protein's 3D structure. As shown in (Fig. 2), except HSP60 all other predicted 3D models are similar in structure. In the 3D structure of proteins, the properties observed were alpha helix (pink color) and show as coils of ribbon, beta sheets (blue color), loop are shown as strings, C and N terminals and hydrogen bonds.

3D structure of proteins is determined by its amino acid sequences and theses sequences are important for the better understanding of the structure and function of proteins (19).

CONCLUSION

The main objective of this study was to use variety of bioinformatics tools and with the help of those tools predict the secondary structure, physiochemical properties, transmembrane domain and 3D structure of the specific proteins which has been differentially expressed in sera of patients suffering from BC. The similarity between physicochemical properties of two or more proteins can help the biologists to further work on that proteins specifically for Breast cancer diagnostics. In present, scientists are working continuously to better understand the relationship between amino acids sequence and predicting protein's 3D structure which will help them for the better analysis of those proteins in future.

ACKNOWLEDGEMENT

I am very thankful to my supervisors Dr. Mohammad Zahid Mustafa, Assistant Professor in the Department of Biotechnology and Molecular Biology, CASVAB (University of Balochistan,) Quetta, for his constant support, guidance, encouragement and strong motivation. Also, I want to thanks my family and friends.

References:

1. Heo GE, Kang KY, Song M, Lee J-H. Analyzing the field of bioinformatics with the multi-faceted topic modeling technique. BMC bioinformatics. 2017;18(7):251.

2. Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucleic acids research. 2019;47(D1): D520-D8.

3. Giordano SH. Breast cancer in men. New England Journal of Medicine. 2018;378(24):2311-20.

4. WHO WHO. Facts Sheet Cancer. WHO; 2018.

5. Rahman SA, Al–Marzouki A, Otim M, Khayat NEHK, Yousef R, Rahman P. Awareness about breast cancer and breast self-examination among female students at the university of sharjah: a cross-sectional study. Asian Pacific Journal of Cancer Prevention: APJCP. 2019;20(6):1901.

6. Zaheer S, Shah N, Maqbool SA, Soomro NM. Estimates of past and future time trends in age-specific breast cancer incidence among women in Karachi, Pakistan: 2004–2025. BMC public health. 2019;19(1):1001.

7. Duffy M, Shering S, Sherry F, McDermott E, O'higgins N. CA 15–3: a prognostic marker in breast cancer. The International journal of biological markers. 2000;15(4):330-3.

8. Piura E, Piura B. Autoantibodies to tumor-associated antigens in breast carcinoma. Journal of oncology. 2010;2010.

9. Bhardwaj N, Stahelin RV, Langlois RE, Cho W, Lu H. Structural bioinformatics prediction of membrane-binding proteins. Journal of molecular biology. 2006;359(2):486-95.

10. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research. 2003;31(13):3784-8.

11. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000;16(4):404-5.

12. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. Protein Data Bank (PDB): the single global macromolecular structure archive. Protein Crystallography: Springer; 2017. p. 627-41.

13. Wulandari D. SDS-PAGE. 2017.

14. Nowakowski AB, Wobig WJ, Petering DH. Native SDS-PAGE: high resolution electrophoretic separation of proteins with retention of native properties including bound metal ions. Metallomics. 2014;6(5):1068-78.

15. Westermeier R. Sensitive, quantitative, and fast modifications for Coomassie Blue staining of polyacrylamide gels. Proteomics. 2006;6(S2 1):61-4.

16. Brunelle JL, Green R. Coomassie blue staining. Methods in enzymology. 541: Elsevier; 2014. p. 161-7.

17. Khaliq M, Tanveer Z. Bioinformatics analysis of serologic proteins of thyroid cancer patients separated by SDS-PAGE. Pak-Euro Journal of Medical and Life Sciences. 2020;3(2):45-9.

18. Singh S, Mittal A. Transmembrane domain lengths serve as signatures of organismal complexity and viral transport mechanisms. Scientific reports. 2016;6(1):1-17.

19. Hu J, Qin H, Li C, Sharma M, Cross TA, Gao FP. Structural biology of transmembrane domains: efficient production and characterization of transmembrane peptides by NMR. Protein Science. 2007;16(10):2153-65.