ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Statistical Studies of Different Cancer Causing Protein Sequences.

Smruti Mayipanda, Divya Srivastava, Santosh Kumar Mishra*, and Priya Ranjan Kumar.

Department of Biotechnology, IMS Engineering College, Ghaziabad, Uttar Pradesh, India.

ABSTRACT

Cancer is both a disease and disorder which is caused by genetic and non-genetic mechanisms that affect the biological and chemical function or expression of normal genes. The effect of physical parameters on carcinogenicity of proteins were calculated and observed and analyzed. After analyzing these data this have been observed that, mean pI value, net surface charge and electrostatic potential of the proteins were considered and effect of these physical factors were observed on for carcinogenicity of the proteins. The objective of this study was to find out the effects of various physical parameters of cancer causing proteins and its effect on the carcinogenicity using statistical study of these proteins sequences The result of study reveals that carcinogenicity increased with decrease in net surface charge and mean pI value and increase in energy conformation whereas length or number of amino acids, molecular weight and hydrophobic nature has less impact or plays no significant role in causing cancer.

Keywords: Carcinogenicity, Homology, Statistical tests, iso-electric point,

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Cancer is a class of diseases involving abnormal cell growth with potential to spread other body parts**.** There are over 100 different types of cancer, and each is classified by the type of cell that is initially affected. According to the report of World Cancer Research American Institute for Cancer Research, Food, Nutrition and the Prevention of Cancer in 1997, meaty, high protein diets were linked with some types of cancer. With the advent of many diseases, various factors causing the disease also came into consideration. Cancer is one of such lethal disease for which there are numerous factors, one of which is protein [1]. It is a complex disease which is caused by both genetic and non-genetic mechanisms and alterations in DNA that affect the biological and chemical function or expression of normal genes. Generally cancer is a multi-step process which results in uncontrolled growth- from normal cell to carcinoma cell [2]. Cancer causing genes are of three distinct groups: proto-oncogenes, tumor suppressor genes and stability genes. In today's scenario, cervical cancer, breast cancer and prostate cancer are the most causing disease which accounts for nearly 1 of every 4 deaths [3]. The computational era of cancer research has started to identify and detect the transcriptomic and genetic difference between normal and cancerous cells. Bioinformatics and statistical informatics aims to develop statistical methods and efficient computational algorithms to overcome quantitative challenges in cancer research for calculation of different parameters of carcinogenic proteins and its activity (National Cancer Institute, U.S. Department of Health and Human Services). Statistical methods are important because they help to minimize errors in the application of trial results. It provides useful quantitative descriptors for summarizing our data. Various statistical tools are used such as Chi-square test, t-test, Mann-Whitney test, Bayesian approach, etc [4]. It also includes more complex statistics such as the correlation between related measurements, the slope of a linear regression and the odds ratio for mortality under differing conditions. These can all be useful for interpreting our data, making informed conclusions, and constructing hypothesis for future studies [5]. Statistics also informs us about the accuracy of the very estimates that we've made. Data analysis usually involves multiple steps, including quality control of raw reads, reads mapping to the reference genome, variant calling, and annotation and prioritization of potentially cancer related variants [6]. Although many software and pipelines are available there is an urgent need of evaluations of statistical methods and tools based on both simulated and benchmark datasets so that users can make appropriate choices for their own data analysis [7]. With the help of statistical analysis it was also known that how physical parameters is linked with chemical parameters of proteins. Changes brought by physical parameters such as negative and positive charge ratio, surface charge, isoelectric point, etc directly cause the chemical modification in protein structure via changing the 3D structure of proteins[8].

MATERIALS AND METHODS

Screening and selection of cancer causing proteins

Three classes of cancer were taken into consideration on which recent research was going on, i.e. cervical cancer, breast cancer and prostate cancer. Cancer causing proteins were searched from literatures available on PUBMED, PMC, OMIM and PDB database from NCBI. Their structures and statistical results given by other researchers were also analyzed from literatures.

Sequence retrieval, Data refinement and selection of suitable protein sequences

Proteins sequences of cancer collected from different database were retrieved by using their names and accession number from databases i.e. NCBI.

Calculation of different physical parameters by using different online tools

Protein sequences were characterized by various tools after sequence retrieval of protein sequences. The various physical parameters found out with the help of ProtParam tool were no. of amino acids, molecular weight, pI, negatively charged residues, positively charged residues, net surface charge ratio and hydropathicity. t-test and chi-square online test were also used to compare these different physical parameters on the basis of accepting and rejecting the hypothesis.

Calculation of different physical parameters based upon homology modeling

Other than these physical parameters, electrostatic potential and bond energy of proteins were also calculated by using Automated Swiss-PDB Viewer. Homology modeling structure prediction was carried out by Geno3D tool which also calculated the energy of the models generated by homology modeling of the protein sequence with the selected templates and the mean deviation. Protein structure visualization was carried out by RasMol tool which was used for the depiction and exploration of biological macromolecule structures. After these, statistical analysis of all these data was carried out in which the physical parameters were compared with each other with the help of t-test and chi-square test on the basis of assumptions and hypothesis.

RESULTS AND DISCUSSIONS

After statistical analysis of selected protein sequences data of cervical cancer, breast cancer and prostate cancer have been tabulated.

Mean pI value and net surface charge of selected protein sequences

TYPES OF CANCER	MEAN pI VALUE	NET SURFACE CHARGE
CERVICAL CANCER		
High risk proteins	6.61	Positive
Moderate risk proteins	**7.735**	**Positive**
Low risk proteins	4.515	Negative
BREAST CANCER		
High risk proteins	**5.645**	**Negative**
Moderate risk proteins	5.31	Negative
Low risk proteins	6.77	Positive
PROSTATE CANCER		
High risk proteins	5.205	Negative
Moderate risk proteins	**7.115**	**Positive**
Low risk proteins	7.905	Positive

From the above results, this has been observed that net surface charge and mean isoelectric point were significant physical parameters which affects the carcinogenicity of proteins. This was because if there was high pI (low pH) then the carboxyl group accepted a proton and became uncharged, so that the overall charge on the molecule was positive and this resulted in positive net surface charge. Similarly, at low pI (high pH) the amino group lose its proton and became uncharged, thus the overall charge on the molecule was negative and this resulted in negative net surface charge. The net surface charge and mean pI value calculation suggested that cervical cancer and prostate cancer causing proteins showed positive surface charge and higher mean pI value and breast cancer causing proteins showed negative surface charge and lower mean pI value. This was consistent with the finding that breast cancer causing proteins were more carcinogenic than cervical cancer and prostate cancer causing proteins because of lower mean pI value and negative surface charge ratio.

Result of t-test

TYPES OF CANCER AND TEST PERFORMED	HIGH RISK PROTEINS AND MODERATE RISK PROTEINS	MODERATE RISK PROTEINS AND LOW RISK PROTEINS	LOW RISK PROTEINS AND HIGH RISK PROTEINS
CERVICAL CANCER			
t-test	2.2105	2.5194	14.79
P-value	0.27	0.2405	0.0430
BREAST CANCER			
t-test	4.6150	5.18	2.54
P-value	0.1358	0.1213	0.2384
PROSTATE CANCER			
t-test	6.2135	4.2027	2.6030
P-value	0.1016	0.1487	0.2335

In case of cervical cancer, high and moderate risk proteins showed positive surface charge and possess higher mean pI value (for high risk proteins= 6.61 and moderate risk proteins= 7.735) whereas in case

of breast cancer, high and moderate risk proteins showed negative surface charge and possess lower mean pI value (for high risk proteins= 5.645 and moderate risk proteins= 5.31) In case of prostate cancer, moderate and low risk proteins showed same positive charge and possess higher mean pI value (for moderate risk proteins= 7.115 and low risk proteins= 7.905). This supported the hypothesis that breast cancer causing proteins were more harmful and carcinogenic than cervical and prostate cancer causing proteins.

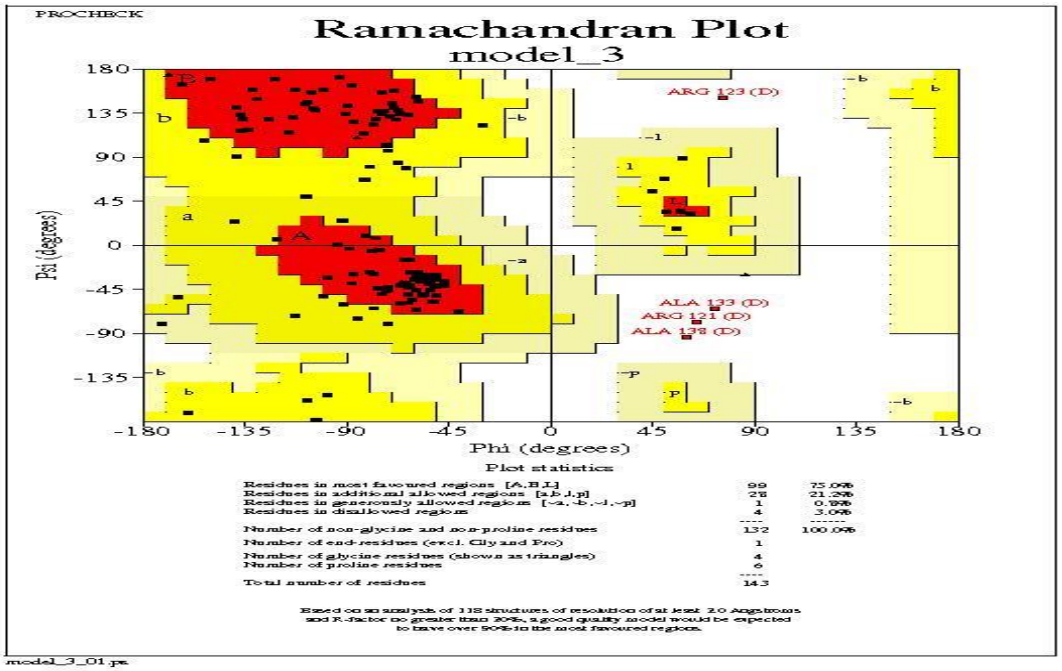
Results of chi-square test

TYPES OF CANCER AND PARAMETERS	HIGH RISK PROTEINS AND MODERATE RISK PROTEINS	MODERATE RISK PROTEINS AND LOW RISK PROTEINS	LOW RISK PROTEINS AND HIGH RISK PROTEINS
CERVICAL CANCER			
χ^2 *test*	6.131	12.622	4.325
P-value	0.0145	0.038	0.056
BREAST CANCER			
χ^2 *test*	0.796	2.769	6.87
P-value	0.3722	0.116	0.0087
PROSTATE CANCER			
χ^2 *test*	2.724	2.685	10.702
P-value	0.112	0.101	0.001

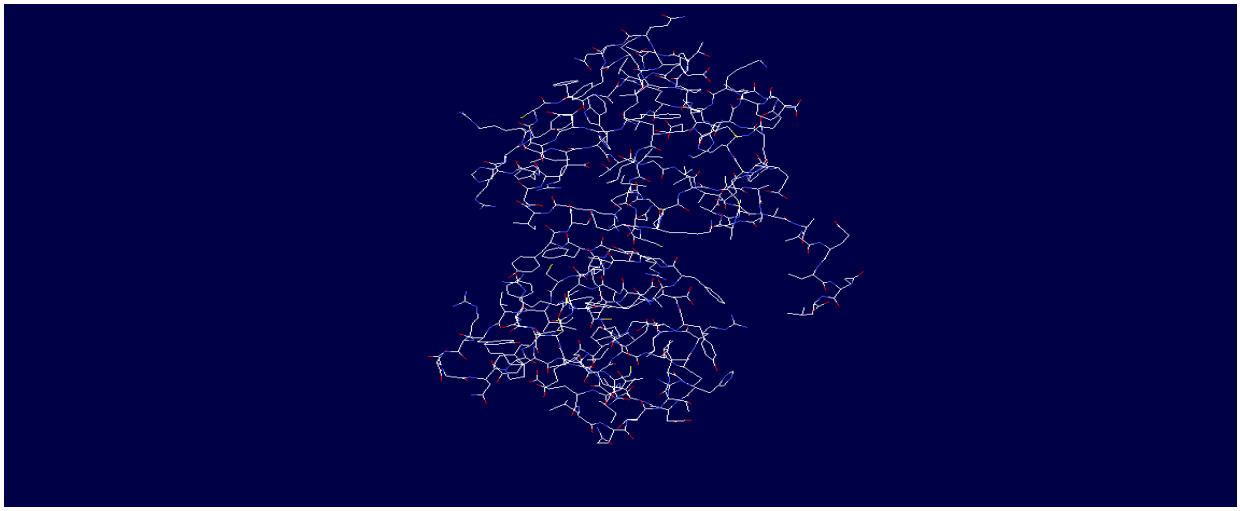
The same results was given by this test also and proved by statistically finding out χ^2 values of different proteins and then from χ^2 table finding the probabilities at 5% level of significance. The significance of the χ^2 values was tested according to the significance of the probabilities.

Results of homology modeling

The homology modeling results from Geno3D indicated that in case of cervical cancer, with increasing value of energy (-5983.46 kcal/mol), there was an increase in carcinogenicity of the cancer causing proteins with decrease in positive surface charge, whereas in case of breast and prostate cancer, the condition was little opposite, i.e., there was an increase in value of energy (for breast cancer= 7989.06 kcal/mol and for prostate cancer= 9176.72 kcal/mol) with increase in carcinogenicity of the cancer causing proteins and increase in their negative surface charge.



ISSN: 0975-8585



Electrostatic potential calculations from Swiss-PDB Viewer tool for E6 oncoprotein responsible for cervical cancer

The values of mean deviation as obtained from Geno3D indicated that in case of cervical cancer, the mean deviation decreased with increase in carcinogenicity of proteins, i.e., low risk proteins have higher mean deviation, whereas in case of breast and prostate cancer, the mean deviation increased with increase in carcinogenicity of proteins, i.e., high risk proteins have higher mean deviation.

During the research work this have been also find out the physical factors i.e. isoelectric point, net surface charge ratio, hydropathicity, molecular weight, number of amino acids, etc which affects the carcinogenicity of various cancer causing proteins and on the basis of homology modeling this has been also observed that bond energy, structure of proteins, electrostatic and hydrogen bonds played a significant role in affecting its carcinogenicity.

CONCLUSION

In the present study we observed the effect of physical factors on carcinogenicity of proteins and how these factors play significant role in the cancer causing proteins on the basis of statistical analysis. In case of cervical cancer, high and moderate risk proteins showed positive surface charge and high mean pI value and thus, these proteins were carcinogenic. However in breast cancer, high and moderate risk proteins showed negative surface charge and low mean pI value and also found highly carcinogenic. In prostate cancer, moderate and low risk proteins showed positive charge and high mean pI value and thus, these were also carcinogenic. This was confirmed by t-test and Chi-square test which showed that two proteins (high or moderate or low) were different or similar from each other on the basis of surface charge in accordance with the hypothesis (rejected or accepted) and how far they were responsible for cancer with respect to net surface charge. Electrostatic potential reflected the potential energy of proton at a particular location which indicated that the increase in energy and mean deviation leads to higher energy conformation of proteins and proved to be highly carcinogenic even in mutated form. Hence, carcinogenicity increased with decrease in surface charge ratio and mean pI value and increase in energy conformation whereas length or number of amino acids, molecular weight and hydropathicity have less or no role to play in causing cancer.

REFERENCES

[1] Li Y Choi PS, Casey SC, Dill DL, Felsher DW. Cancer Cell. 2014; 26:262-72.
[2] Hanahan D, Weinberg RA. Cell 2000; 100:57-70.
[3] Vogelstein B, Kinzler KW. Nat Med 2004; 10:789-799.
[4] Simon J Furney, Desmond G Higgins, Christos A Ouzounis , Núria López Bigas.BMC Genomics 2006; 7:1-2.
[5] Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS. Nature medicine 2001;7:673-679.

[6] Van't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ. Nature 2002; 415 :530-536.
[7] Jiang H, An L, Baladandayuthapani V, Auer .L. Cancer informatics 2014;13:1-3
[8] Besa Smith Ames, Bruce N Gold, Lois Swirsky. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2000; 447: 3–13.