In Silico Approach for Prediction of the Structural and Functional Impacts of 184(T>C) Missense Mutation Identified in *BRCA1* Gene in a Syrian Breast Cancer Patient

Husam Khalil♦

Al-Rasheed International University for Science and Technology (RU), Damascus, Syria

Abstract

Background: The classification of genetic variations depending on their clinical impacts is highly relevant for clinical decision-making. Therefore, predicting the effects of missense mutations using in silico tools has become a frequently employed approach. The objective of this study was to analyze the impacts of a previously detected *BRCA1* missense mutation using an in silico prediction tool in the context of invasive breast cancer.

Methods: In this bioinformatics study, application of the in silico combination of tools Phyre2 to 184(T>C), a *BRCA1* missense mutation previously characterized by Khalil et al. in 2018, we used to predict its clinical impacts.

Results: Incidence of the missense mutation caused a disorder in the zinc binding RING finger functional domain of *BRCA1* protein. This incidence was considered to be a major contributor to the interaction of this protein with other proteins in signal transduction pathway in the mechanism of cellular response to DNA damage. The functional analysis also revealed that the detected missense mutation might significantly affect the function of both the protein and phenotype of the living organism.

Conclusion: In silico prediction confirmed the detrimental impact of the identified missense mutation in the exon 2 of *BRCA1* gene on both the structural and functional properties of the generated *BRCA1* protein.

Keywords: In silico, Missense, 184(T>C), Syria, Phyre2

♦**Corresponding Author:**
Husam Khalil, PhD
Al-Rasheed International University for Science and Technology (RU), Damascus, Syria
Email: husam.khalil86@gmail.com

Introduction

Sequencing of a high-risk cancer susceptibility gene such as *BRCA1* or *BRCA2* may reveal that a patient carries a clearly pathogenic variant. Most pathogenic variants in these genes are nonsense variants, small insertion, or deletion variants (indels) that create a frameshift, larger gene rearrangements, variants generating a severe splicing aberration, or severely dysfunctional missense

Home sapiens breast and ovarian cancer susceptibility _(BRCA1)_ mRNA, complete cds.

Figure 1. Sequence of the target region in _BRCA1_ gene and position of the identified mutation 184(T>C).

substitutions.[1] Depending on their clinical impact, the classification of genetic variations is highly relevant for clinical decision-making. For individuals at risk for breast cancer, undergoing prophylactic surgery is limited to carriers of pathogenic mutations in relevant risk genes. In addition, it is important to know the _BRCA1_/2 mutation status of patients affected by breast cancer because it determines the therapeutic response and medication alternative (Such as PARP inhibitors).[2, 3]

Prior to family history screening, it is essential to confirm the pathogenicity of the variant. While nonsense and frameshift mutations causing the premature termination of protein production are usually pathogenic, missense mutations are commonly benign.[4] The majority of detected mutations are classified as variants of unknown clinical significance (VUS).[5] There are also programs that combine the output from several in silico prediction tools and produce a single result; all of these programs have been reported to have a more enhanced performance compared with individual tools.[6, 7] The automated prediction of the effects of missense mutations has become a widely used approach in clinical diagnostics.


Figure 2. Alignment of the amino acid sequence of mutant exon 2 protein product to that of the normal exon and the identified missense mutation [(184(T>C)].

In silico approaches for the classification of missense mutations are dependent upon the assumption that disease-associated missense mutations are characterized by a large difference in the biochemical properties between substituted amino acids[2] and those located at highly conserved genomic regions among species.[3]

In silico tools have their own strengths and weaknesses depending on the algorithm; moreover, in many cases, performance varies depending on the gene and protein.[8,9] The MutPred tool was proven to be the best predictor of variants of genes associated with the limb-girdle muscular dystrophy (LGMD);[10] MAPP and MAPP + PolyPhen-2.1 provided the best combined model for testing the variants of MLH1, MSH2, MSH6, and PMS2 genes associated with Lynch syndrome, a hereditary form of colon cancer;[11] the SIFT had the most optimal performance regarding the variants of the UGT1A1 gene associated with Crigler-Najjar syndrome;[12] the Align GVGD in silico tool was reported as the best for analyzing the variants of *BRCA1*/2 genes.[13]

Align-GVGD aligns multiple sequences to compute the value of two parameters, namely a biochemical distance score (extension of the pairwise Grantham difference, GD) and a conservation score (Grantham variation, GV) for each alignment representing a substitution.

Based on the computed values of GD and GV, substitutions are categorized into seven classes [0,15,25,35,45,55,65] from least likely to interfere with function to most likely to interfere with
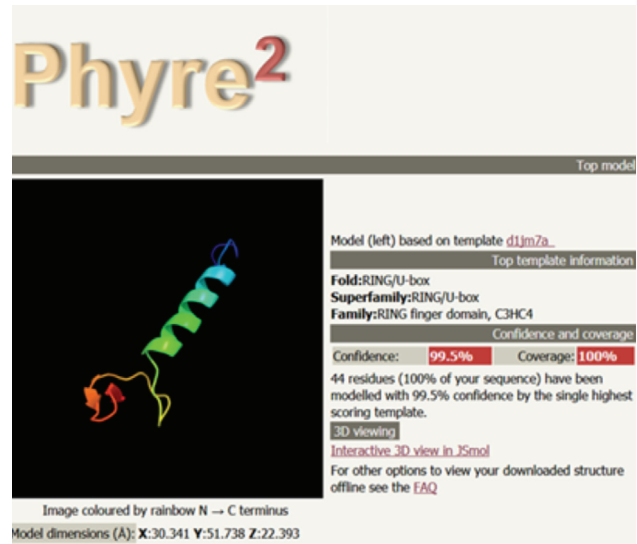
Figure 3. Predicted 3D structural model of the normal protein product of exons 2 and 3 in *BRCA1* gene.

function.[14, 15] As a purely sequence-based prediction tool, SIFT classifies non-synonymous single nucleotide polymorphisms (nsSNPs) based on the evolutionary conservation of amino acids within protein families. A scaled probability for each AA substitution to occur (SIFT score) is computed at each position of an alignment. A SIFT score of a substituted AA below a threshold of 0.05 means that the missense variant is predicted to have a harmful effect on protein function.[16]

MutationTaster2 categorizes variants into either disease causing or polymorphism based on characteristics such as evolutionary conservation degree and splice site predictions.[17] PolyPhen-2 (Polymorphism Phenotyping v2) uses eight
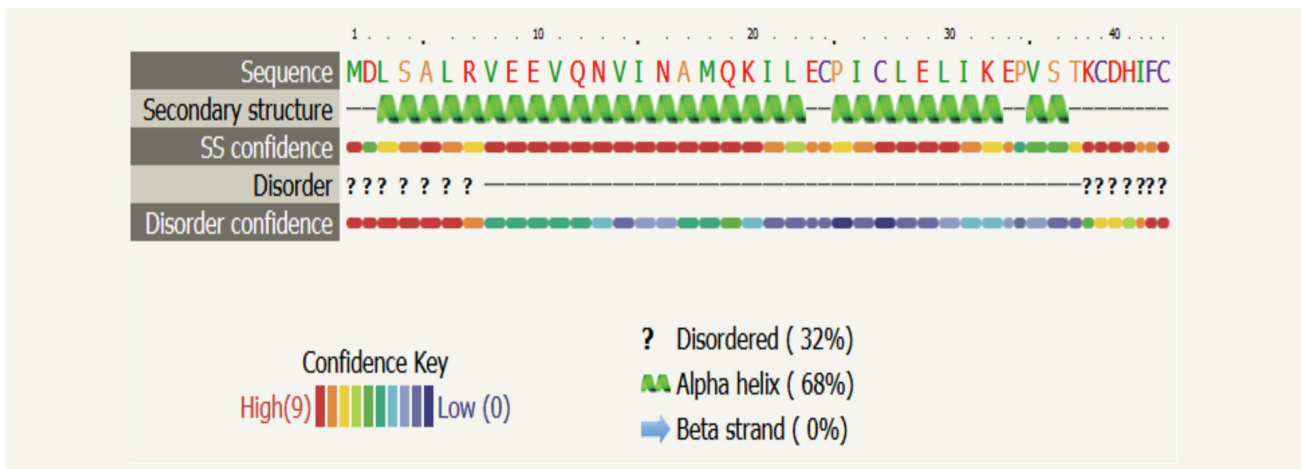
Figure 4. Predicted secondary structure of the normal protein product encoded by exons 2 and 3 of *BRCA1* gene.

sequence-based and three structural features as the input for a naïve Bayes classifier. PolyPhen-2 is considered only in cases where a 3D structure is known for the protein of interest to predict whether or not a missense mutation is situated in a structurally important/functional site in the protein.[18] Santos et al. highlighted the importance of integrating the clinical and familial data with theoretical models that are not able to characterize variants alone.[19]

Phyre[2] is a web-based service for protein structure prediction, which is free for non-commercial use. Over 1500 bioinformatics studies have used this service worldwide.[20] Phyre[2] predicts the 3D structure of the protein via homology modeling technique. In this approach, the sequence of the target protein can be modeled with an acceptable accuracy on a very remotely related sequence of known structure (the template) provided the sequence alignment confirms the relationship between target and template. Currently the most accurate methods for aligning and detecting distantly-related sequences depend on profiles that capture the mutational tendency of each position in an amino acid sequence based on the observed mutations in the related sequences.[21,22] Typically, the amino acid sequences of a representative set of all known three-dimensional protein structures are collected and processed by scanning against a large protein sequence database. The result is a database of profiles, each belonging to a known 3D structure. The sequence of interest is similarly processed
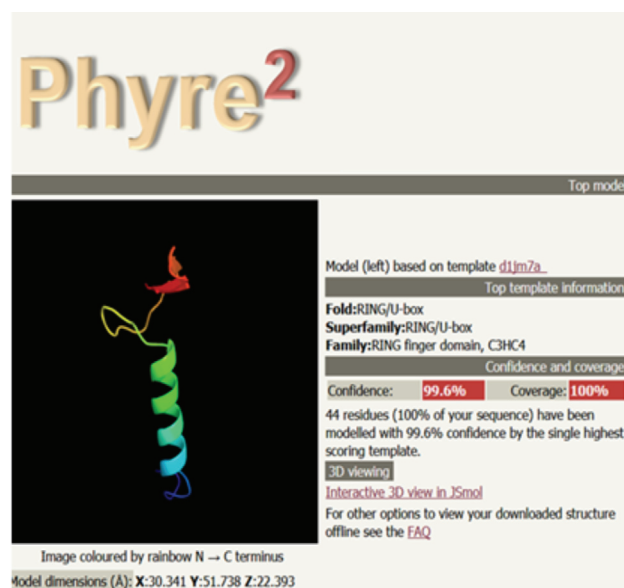
Figure 5. Predicted 3D structural model of the protein product encoded by mutant exon 2 and normal exon 3 in *BRCA1* gene.

to form a profile which is then scanned against the database of profiles using profile-profile alignment techniques. These alignments further consider the patterns of predicted or known secondary structure elements and can be scored using many statistical models.[22]

A group of Syrian women diagnosed with an invasive breast cancer participated in our previous study; we identified a missense mutation in the exon 2 of *BRCA1* gene [184(T>C)] in one of the breast cancer patients. In the present study, we aimed to predict the structural and functional impacts of the identified missense mutation using an in silico prediction.
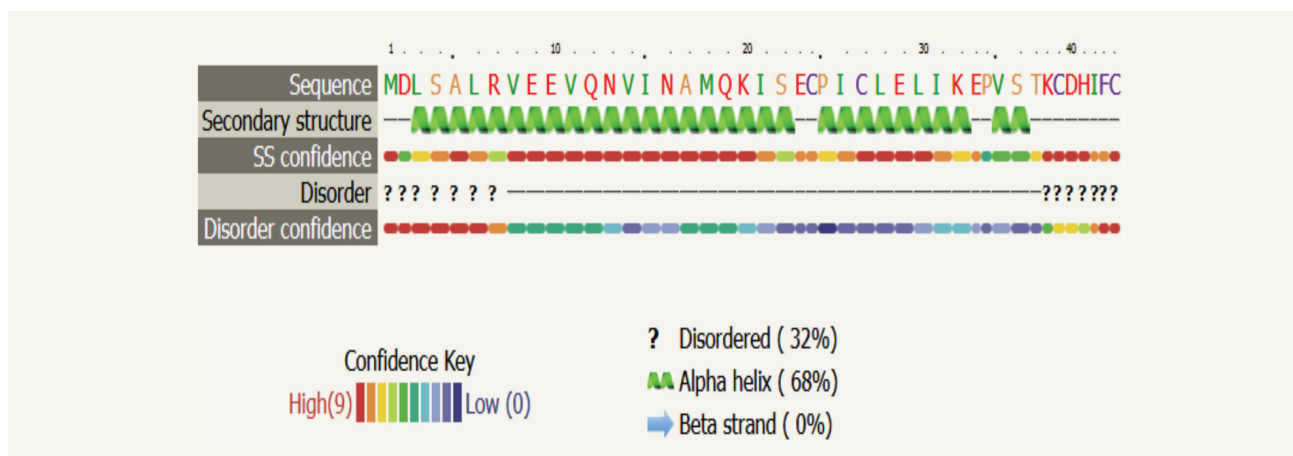
Figure 6. Predicted secondary structure of the protein product encoded by mutant exon 2 and normal exon 3.
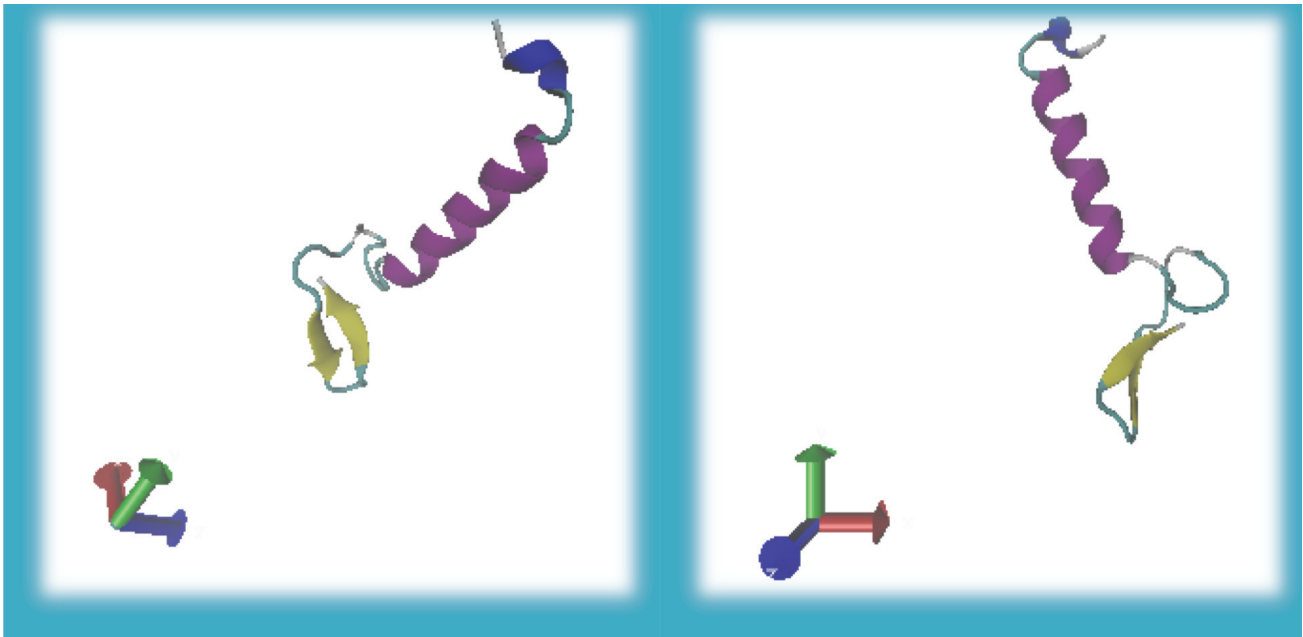
Figure 7. The 3D model of the normal protein product visualized by VMD 1.9.3 OpenGL software (colored according to secondary structure).

Materials and Methods

The in silico combination of tools Phyre[2] (www.sbg.bio.ic.ac.uk/phyre[2]/) to 184(T>C), a *BRCA1* missense mutation previously characterized by Khalil et al. in 2018,[23] was applied to predict its clinical impacts. The prediction was performed using the following procedure:

• First, an alignment between the obtained sequence and the target region containing the exon 2 was performed to determine the matched region between the obtained sequence and the exon 2, identify the position of the detected mutation, and determine whether it occurred inside or outside the exon of interest.

• After ensuring the occurrence of the mutation in a protein coding region, the location of the detected mutation in the reference full sequence
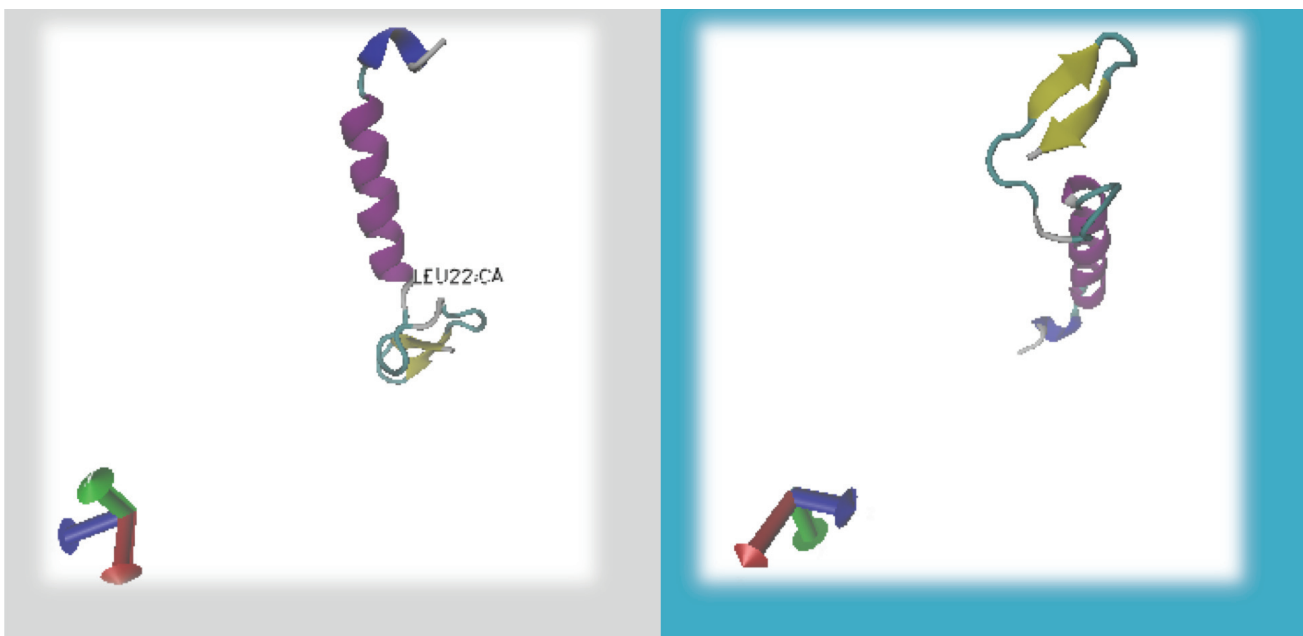
Figure 8. ERRAT analysis of the entire structure of the protein encoded by exons 2 and 3 in *BRCA1* gene.
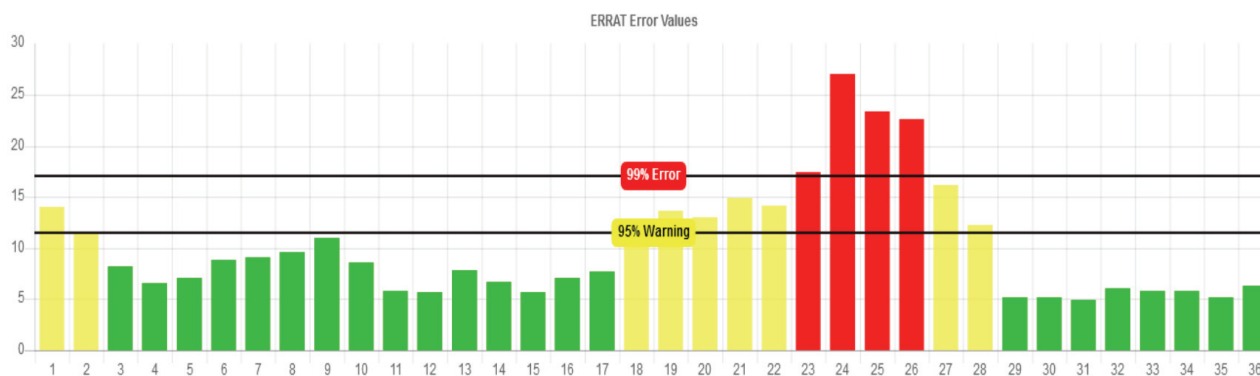
Figure 9. The values of phi/psi combinations related to all residues of the modeled protein product of exons 2 and 3 viewed by Ramachandran plot.

of exon 2 was set, which was obtained from GenBank database. Geneious software R9 (Biomatters Ltd.) determined the most probable amino acid sequence of exon 2 protein product.

• Finally, both sequences of natural and mutant protein products of the exon 2 was entered into Phyre[2] server to predict their 3D models and secondary structures involving functional domains. The structure models achieving the highest match scores with the entered protein sequences were the most probable models representative of those sequences.
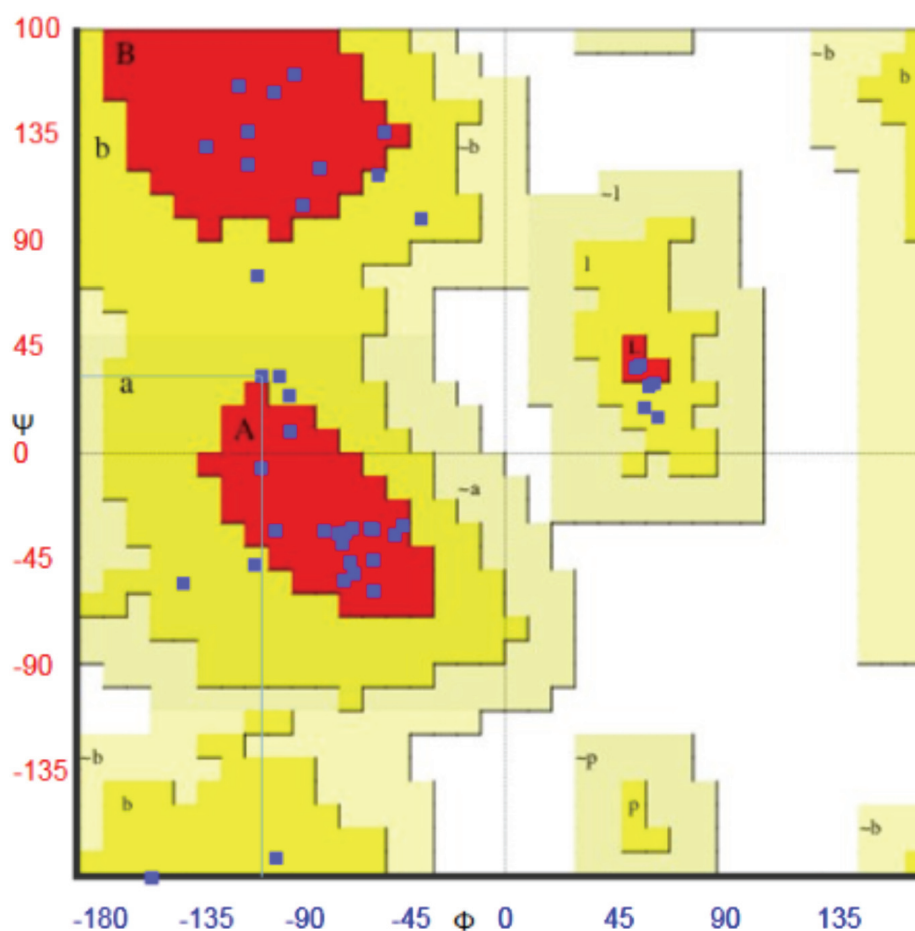
Figure 10. The results of analyzing the PDB-structure of protein product of exons 2 and 3 using PROVE (involving the calculated statistical atomic Z-scores, Z-score mean, Z-score standard deviation, and Z-score RMS). The percentage of outlier atoms from the analyzed structure is 4.0% (Warning).

Phyre[2] Investigator investigated the possible functional effects of the detected missense mutation; such advanced alternative allowed us to perform a group of protein functional analyses including:

• Conservation analysis to 1) analyze the conservation of the amino acid sequence of protein depending on a bioinformatics approach called Jensen-Shannon variation, and 2) classify amino acid residues according to their relative importance regarding the proper protein structure and function.

• PI-Site interface residues analysis using PI-Site protein-protein interface database.

• Pocket detection by the use of fpocket2 program. It has been frequently found that large pockets form locations for the protein active sites.

• Mutational sensitivity analysis which predicts the effects of each of the twenty possible missense mutations in a specific position in the protein sequence on both the protein function and phenotype of the organism. SuSPect tool conducted that analysis.[21]

Results

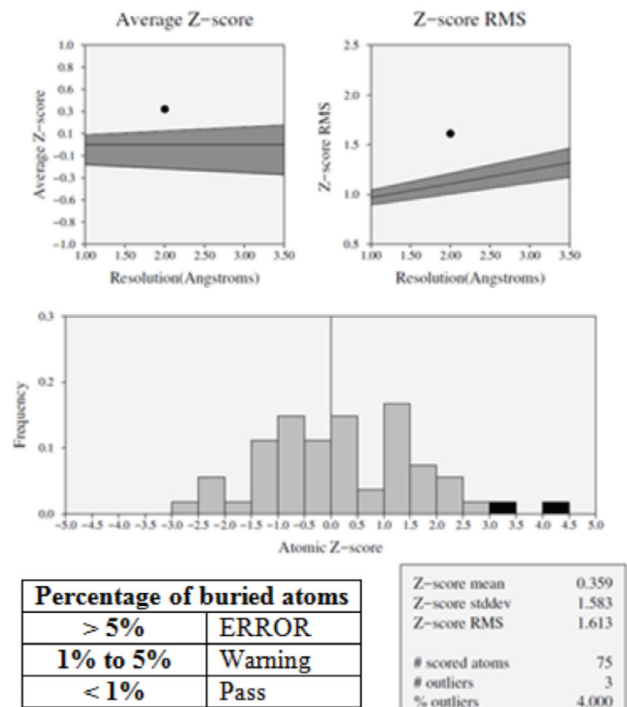Figure 1 shows the nucleotide sequence of the

Figure 11. Conservation analysis of the Leucine residue in the position of substitution (22).

target region (the shaded region) spanning from 43,123,921 to 43,124,178, involving the exon 2 in *BRCA1* gene and position of the missense mutation
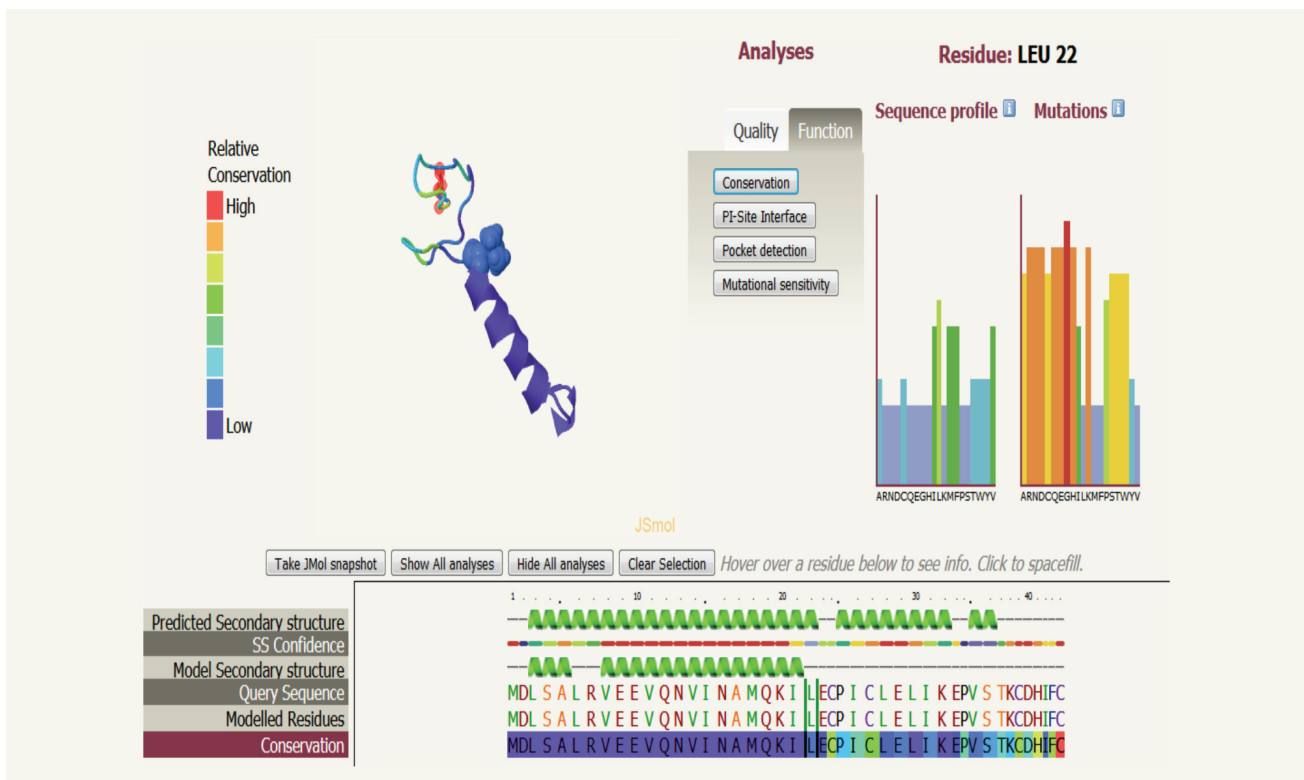

Figure 12. PI-Site interface residues analysis of the replaced leucine residue.

detected in the exon 2 referred to by an arrow.

We obtained figure 2 through aligning the amino acid sequence of protein product of the mutant exon 2 [(184(T>C)] to that of the normal exon. The length of the wild-type exon 2 is 99 bp and the first 19 nucleotides of that exon form a non-coding region.

Figures 3, 4, 5, and 6 show the predicted protein structures for each of the normal and mutant exons.

Figure 7 shows the 3D model of the normal protein product visualized by VMD 1.9.3 OpenGL software.

Different types of atoms, carbon (C), nitrogen (N), and oxygen (O) are non-randomly distributed in the proteins. This results in six different combinations of pairwise non-covalent interactions (CC, CN, CO, NN, NO, and OO). Errors in model building lead to more randomized distributions of the different atom types. Statistical methods are able to distinguish these distributions from the correct ones. ERRAT uses a quadratic error function to characterize the set of pairwise interactions from nine-residue sliding windows

in a database of 96 reliable protein structures. Analysis of non-covalent interaction patterns from each window can specify the regions of candidate protein structures that are mistraced or misregistered.[24] Figure 8 shows ERRAT analysis of the entire structure of the protein encoded by exons 2 and 3 in *BRCA1* gene.

Figure 9 shows Ramachandran plot which illustrates the values of phi/psi combinations pertaining to all residues of the modeled protein product of exons 2 and 3. In this figure, red indicates the most favorable regions, yellow represents the allowed region, pale yellow is the generously allowed region, and white refers to the disallowed regions. Phi and psi are the dihedral angles (CO)C----Cα and Cα----N(NH), respectively.

Figure 10 shows the results of analyzing the PDB-structure of the protein product of exons 2 and 3 using PROVE (PROtein Volume Evaluation), which is a program that implements volume-based structure validation procedures to evaluate the quality of protein crystal structures. Standard ranges of atomic and residue volumes
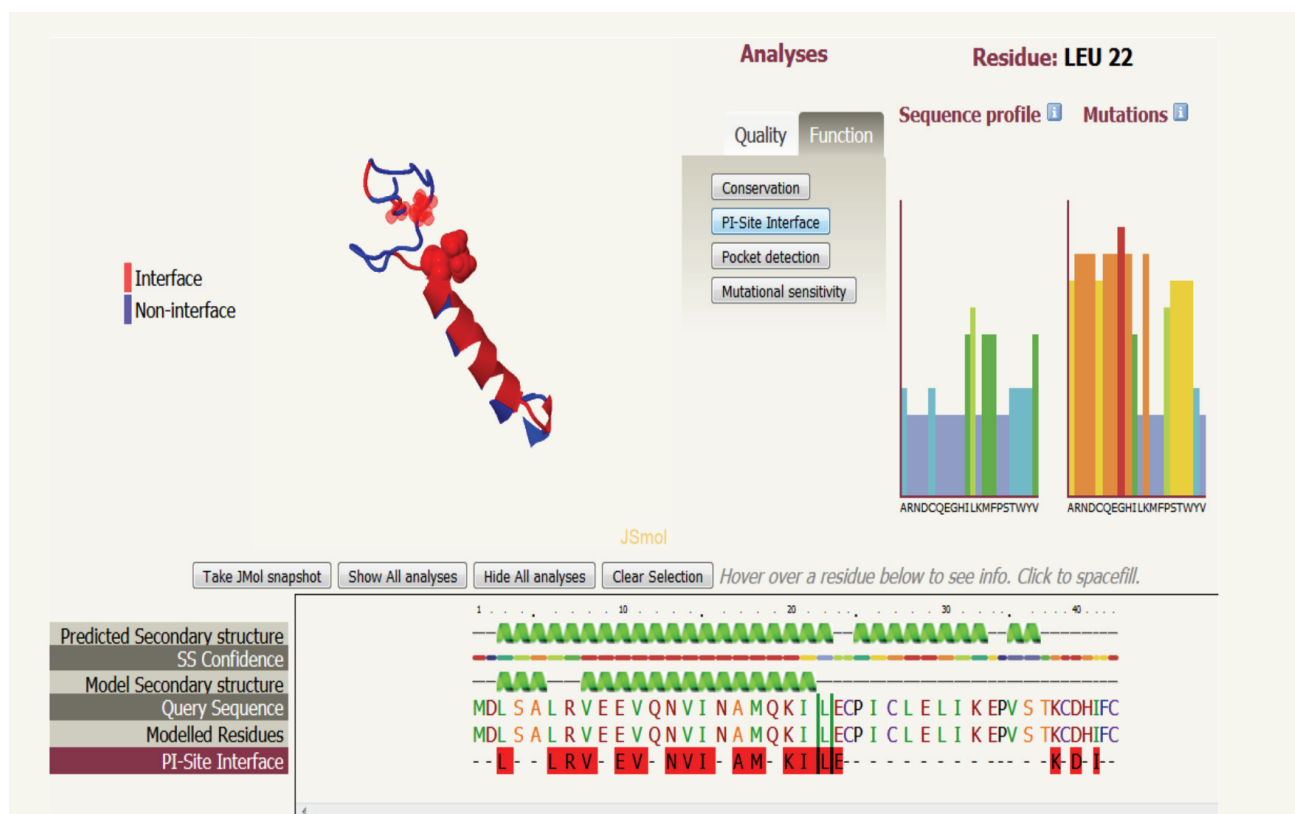

Figure 13. Pocket detection analysis for the predicted localization of the replaced leucine residue.

are calculated in highly resolved and well-refined protein crystal structures. Deviations of the atomic volumes from the standard values are evaluated as the volume Z-scores. The calculated volume Z-score root mean square deviation (Z-score rms) measures the average magnitude of the volume irregularities in the structure. Z-score rms decreases with the increase in the resolution and R-factor, reflecting more accurate models. The normal limits of Z-score rms values are specified in the scored structure. Structures whose Z-score rms exceeds these limits are considered as outliers. As revealed by other analyses, such structures also exhibit unusual stereochemistry.[25]

Figures 11, 12, 13, and 14 show the results obtained from the functional analyses performed by Phyre[2] Investigator tools

Figure 11 shows the conservation analysis of the Leucine residue in the position of substitution (22).

Figure 12 shows PI-Site interface residues analysis of the replaced Leucine residue.

Figure 13 shows the Pocket detection analysis for the predicted localization of the replaced Leucine residue.

Figure 14 shows the result of mutational sensitivity analysis.

Discussion

The alignment of the amino acid sequence related to the protein product of the mutant exon 2 to that of the normal exon showed that a serine residue (S) replaced the leucine (L) in the position 22 of normal protein. This was the result of the missense mutation [(184(T>C)] previously characterized by Khalil et al. in 2018.[23] Substituted amino acids are significantly different regarding their biochemical properties. Leucine is a nonpolar amino acid while serine is polar neutral. Therefore, the detected missense mutation is likely to be associated with the disease.

Exon 2 was shown as a contributor to coding an important functional domain of *BRCA1* protein structure known as amino-terminal zinc binding RING finger domain. Zinc binding RING finger domain consists of two zinc finger-like motifs connected through linking C3HC4 regions. This domain is highly conserved among species. Therefore, its associated disorders may strongly change the binding properties of the protein interfering with its interactions with other proteins involved in the cellular damage repair machinery.[26,27]

Phyre[2] Investigator functional analyses was carried out to investigate the possible functional effects of the detected missense mutation on the protein product; on the other hand, the
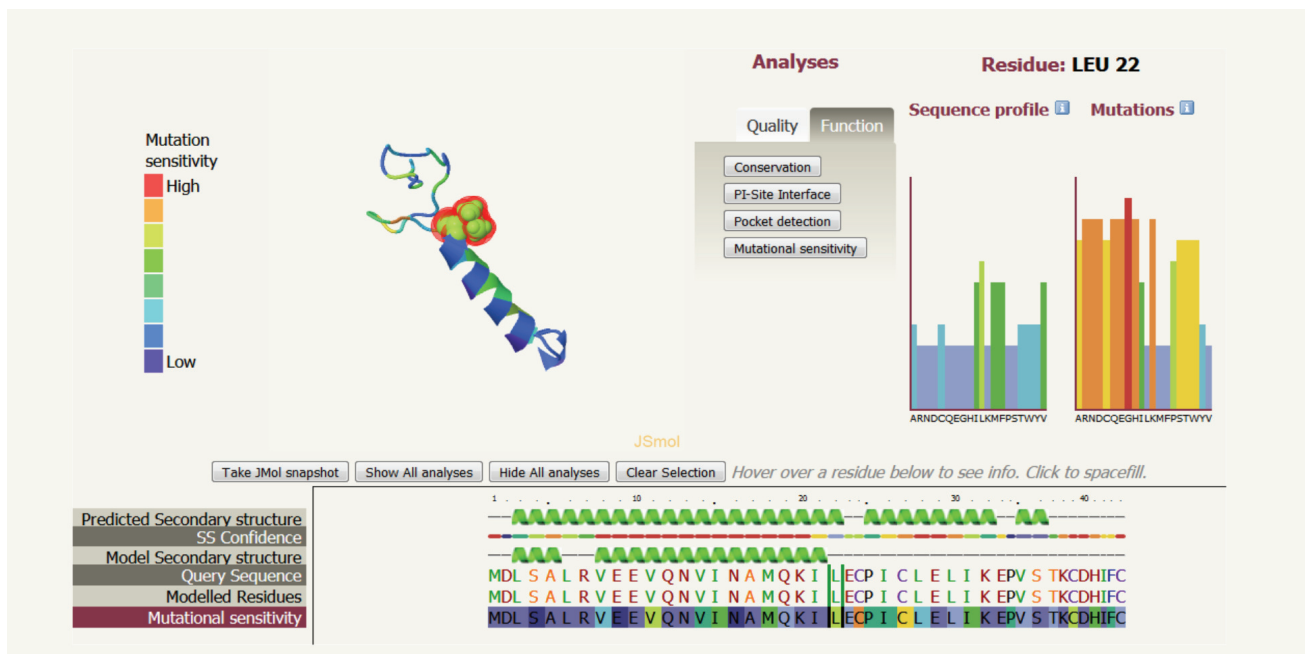

Figure 14. Mutational sensitivity analysis.

conservation analysis showed that the leucine residue in the position of substitution (22) had a low relative conservation and was highly replaceable; it should be noted that conservation degree was determined according to color by checking the relative conservation scale. The PI-Site interface residues analysis of the replaced leucine residue further showed the predicted localization of that residue in the protein-protein interface. However, Pocket detection analysis showed the predicted localization of the replaced leucine residue between two of the largest pockets in the protein sequence (red regions in the visualized structure). Those pockets were frequently considered as the most possible locations for the protein active sites. On the other hand, mutational sensitivity analysis showed that the identified missense mutation (L>S) was highly predicted to affect the protein function and phenotype of the living organism; of note, the mutational sensitivity was determined according to color by checking the mutational sensitivity scale.

Conclusion

In silico prediction confirmed the harmful impact of the identified missense mutation in the exon 2 of *BRCA1* gene on both the structural and functional properties of the generated *BRCA1* protein. Therefore, the occurrence of such mutation may interfere with the important role of *BRCA1* protein as a tumor suppressor protein through disrupting the protein's ability to interact with other proteins involved in the nuclear complex responsible for the regulation of transcription, cell cycle control, and DNA damage repair.

Conflict of Interest

None declared.

References

1. Vallée MP, Di Sera TL, Nix DA, Paquette AM, Parsons MT, Bell R, et al. Adding in silico assessment of potential splice aberration to the integrated evaluation of BRCA gene unclassified variants. *Hum Mutat.* 2016;37(7):627-39. doi: 10.1002/humu.22973.

2. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben Eric D, Zendulka J, et al. Predict SNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol.* 2014;10(1). https://doi.org/10.1371/journal.pcbi.1003440.

3. Ernst C, Hahnen E, Engel C, Nothnagel M, Weber J, Schmutzler RK, et al. Performance of in silico prediction tools for the classification of rare *BRCA1*/2 missense variants in clinical diagnostics. *BMC Med Genomic.* 2018;11(1):35. doi: 10.1186/s12920-018-0353-y.

4. Kerr I, Cox H, Moyes K, Evans B, Burdett B, van Kan A, et al. Assessment of in silico protein sequence analysis in the clinical classification of variants in cancer risk genes. *J Community Genet.* 2017;8(2): 87-95. doi: 10.1007/s12687-016-0289-x.

5. Sadowski C, Kohlstedt D, Meisel C, Keller K, Becker K, Mackenroth L, et al. *BRCA1*/2 missense mutations and the value of in-silico analyses. *Eur J Med Genet.* 2017;60(11): 572-7. doi: 10.1016/j.ejmg.2017.08.005.

6. Leong I, Stuckey A, Lai D, Skinner J, Love D. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med Genet.* 2015;16:34. doi: 10.1186/s12881-015-0176-z.

7. Mathe E, Olivier M, Kato S, Ishioka C, Hainaut P, Tavtigian SV, et al. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res.* 2006;34(5):1317-25. doi: 10.1093/nar/gkj518.

8. Olatubosun A, Väliaho J, Härkönen J, Thusberg J, Vihinen M. PON-P: Integrated predictor for pathogenicity of missense variants. *Hum Mutat.* 2012;33(8):1166-74. doi: 10.1002/humu.22102.

9. Pshennikova V, Barashkov N, Romanov G, Teryutin F, Solov'ev A, Gotovtsev N, et al. Comparison of predictive in silico tools on missense variants in GJB2, GJB6, and GJB3 genes associated with autosomal recessive deafness 1A (DFNB1A). *Sci World J.* 2019;1-9. doi: 10.1155/2019/5198931.

10. Rahim F, Galehdari H, Mohammadi-asl J, Saki N. Regression modeling and meta-analysis of diagnostic accuracy of SNP-based pathogenicity detection tools for UGT1A1 gene mutation. *Genet Res Int.* 2013;2013:1-7. doi: 10.1155/2013/546909.

11. Thompson B, Greenblatt M, Vallee M, Herkert J, Tessereau C, Young E, et al. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat.* 2012; 34(1):255-65. doi: 10.1002/humu.22214.

12. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 2011;32(4):358-68. doi: 10.1002/humu.21445.

13. Tutt A, Robson M, Garber J, Domchek S, Audeh M, Weitzel J, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and advanced breast cancer: a proof-of-concept trial. *Lancet.* 2010;376(9737):235-44. doi: 10.1016/S0140-6736(10)60892-6.

14. Walters-Sen L, Hashimoto S, Thrush D, Reshmi S, Gastier-Foster J, Astbury C, et al. Variability in pathogenicity prediction programs: impact on clinical diagnostics. *Mol Genet Genomic Med.* 2014;3(2):99-110. doi: 10.1002/mgg3.116.

15. Tavtigian SV, Samollow PB, Silva Dd, Thomas A. An analysis of unclassified missense substitutions in human *BRCA1*. *Fam Cancer.* 2006;5(1):77–88. doi: 10.1007/s10689-005-2578-0.

16. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073–81. doi: 10.1038/nprot.2009.86.

17. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods.* 2014;11(4):361-2. doi: 10.1038/nmeth.2890.

18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):48-9. doi: 10.1038/nmeth0410-248.

19. Santos C, Peixoto A, Rocha P, Pinto P, Bizarro S, Pinheiro M, et al. Pathogenicity evaluation of *BRCA1* and *BRCA2* unclassified variants identified in Portuguese breast/ovarian cancer families. *J Mol Diagn.* 2014;16(3):324-34. doi: 10.1016/j.jmoldx. 2014.01.005.

20. Kelley L, Mezulis S, Yates C, Wass M, Sternberg M. The Phyre^2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845-58. doi: 10.1038/nprot.2015.053.

21. Kelley L, Sternberg M. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* 2009;4(3):363-71. doi: 10.1038/nprot.2009.2.

22. Bennett-Lovsey RM, Herbert AD, Sternberg ME, Kelley LA. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins.* 2007;70 (3):611-25. doi: 10.1002/prot.21688.

23. Khalil H, Monem F, Al-Quobaili F. Identification of three *BRCA1*/2 mutations and a study of the likelihood of an association with certain characteristics in Syrian familial breast cancer patients. *Middle East J Cancer.* 2018;9(4):274-81.

24. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2(9):1511-9. doi: 10.1002/pro. 5560020916.

25. Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol.* 1996;264(1):121-36. doi: 10.1006/jmbi.1996.0628

26. O'Donovan PJ, Livingston DM. *BRCA1* and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis*. 2010;31(6):961-7. doi: 10.1093/carcin/bgq069.

27. Johnson NC, Kruk PA. *BRCA1* zinc ring finger domain disruption alters caspase response in ovarian surface epithelial cells. *Cancer Cell Int*. 2002;2:7. doi: 10.1186/1475-2867-2-7.