



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume3, Issue2)

Available online at www.ijariit.com

Anxa5 Gene: An Insilco Approach for SNP Analysis Associated With RPL

Nasreen Sultana

Dept. of Biotechnology
Acharya Nagarjuna University
nasreenci@gmail.com

Jeevan Amos. S

Dept. of Biotechnology
Acharya Nagarjuna university
jeevanseelam@gmail.com

Kasturi .K *

Dept. of Biotechnology
Acharya Nagarjuna University
kasturi.is.kondapalli21@gmail.com

Abstract: Miscarriage is the most familiar pregnancy obstacle affecting up to 15% of all clinically recognized pregnancies. Annexin a5 gene is a well-recognized predisposition factor for recurrent spontaneous abortion. Annexin a5 is both Ca^{2+} and phospholipids binding protein that is localized at the surface of the placental syncytiotrophoblast layer and performs a vital role in anticoagulant function in the maternal blood at the intervillous space. It has been recently shown that polymorphisms in the promoter region of the Annexin a5 gene (*anxa5*) are significantly associated with *rpl*. The present study focuses on the comprehensive analysis of the functional and structural impact of all known SNPs of *anxa5* gene by using computational approaches. A comprehensive analysis was performed to filter the most probable mutation that might be associated with *rsa*. It resulted in the identification of SNPs which are most deleterious and disease associated. The study reveals that these two *rs111418524*, *rs41278081* snps have more significant effect on structure and function of the protein.

Keywords: Recurrent Spontaneous Abortions, Annexin A5 Gene, Single Nucleotide Polymorphism, Sift, PolyPhon, Mutant-I, Hope.

INTRODUCTION

Pregnancy is a complex, heterogeneous, biological phenomenon in which the embryo develops into a fetus within the female uterus. Due to various etiological factors, the growing embryo that is unable to survive is expelled from the pregnant mother at different gestational ages and this is referred to as pregnancy loss or abortion. Pregnancy loss occurs in 10 to 15% of all pregnancies, of which 1-2% are recurrent [1]. According to WHO, Recurrent pregnancy loss has been defined as the loss of a fetus weighing $\leq 500g$, at an early period of gestation (20-22weeks) [2]. Recurrent pregnancy loss (RPL) is a complex, multifactorial condition with a polygenic background. [3] There are 30 genes showing different levels of expression between normal and RPL patients [4]

ANXA5 GENE

ANXA5 is a member of the annexin protein family. Annexin A5 (placental anticoagulant protein, encoded by ANXA5 gene) normally occurs on the placental villi, and its expression appears to be decreased in the presence of antiphospholipid antibodies (APL). Annexin A5 functions as an inhibitor of coagulation by its ability to bind to the anionic phospholipids exposed on the surface of platelets. It is abundantly expressed in the normal placenta, and the encoding gene possesses a complex promoter region that is subject to intricate regulation. It has been shown that single-nucleotide polymorphisms (SNPs) in the promoter region of the ANXA5 gene are significantly associated with recurrent pregnancy loss (RPL) and that women with the M2 haplotype have a 2.42- fold higher risk of pregnancy loss than

non-carriers. More recently, it has been shown that an SNP in the 5' untranslated region of the ANXA5 gene (SNP: 302T>G) serves as a major risk determinant of RPL.

SNPs exist throughout the genome and can be divided into several groups. Among the different kinds of SNPs, a non-synonymous SNP in the coding arena of a gene (genetic code) is vital as long as it alters the amino acid form. Consequently, such alterations may have an impact on protein structure, function, and subcellular localization. Although pinpointing the effects of many non-synonymous SNPs using biochemical analysis has been challenging. Computational tools can predict their effect on protein activity and stability. Several research articles stated the effectiveness in identifying the deleterious and disease-associated mutations, thus predicting the pathogenic non-synonymous SNPs in correlation to their functional and structural damaging properties [5, 6, 7, 8]

In this study, *In silico* methods were used to test the possible effect of the missense mutations rs111418524, rs41278081 on the structure and function of ANXA5 (annexin Protein)

MATERIALS AND METHODS

DATA RETRIEVAL

SNP information (SNP rsID, Gene ID) and protein sequence for ANXA5 were obtained from National Centre for Biotechnology Information (NCBI) database dbSNP. (<http://www.ncbi.nlm.nih.gov/projects/SNP>).

SIFT

SIFT tool was used to predict whether rs111418524, rs41278081 affect ANXA5 function, this program was developed by Ng and Henikoff, based on sequence homology [9]. The Concept on which this tool predicts upon is that protein functions are interrelated. Thus any change occurs in an evolutionary conserved region, it is less tolerable than a change in an unconserved region. SIFT algorithm uses a modified version of PSIBLAST [10]. Accordingly, any position with a tolerance index of less than 0.05 is predicted to be deleterious and those with tolerance index greater than or equal to 0.05 are predicted to be tolerated.

POLYPHEN-2 (POLYMORPHISM PHENOTYPING V2)

Polymorphism Phenotyping-2 (PolyPhen-2) tool was used to investigate the possible effect of nsSNPs on protein. The predictions obtained are rooted in physical and evolutionary comparative considerations and naïve Bayes classifier which uses eight sequences and three structures to calculate the deleterious effects of non-synonymous variants [11]. If the scores of PolyPhen-2 are between 0.00 to 0.99, the prediction is to have a benign effect, and if the scores are in the range of 1.0 to 1.99 it could be possibly damaging and if the score is more than 2.0, it will be predicted as probably damaging to the protein [12].

I-MUTANT 2.0

To predict the effect of SNP on protein stability, I-Mutant 2.0 was used. The differences in free energy ($\Delta\Delta G$) between the unfolding Gibbs free energy values of the mutated protein with that of the wild-type (Kcal/mol) are used as a basis for the prediction by I-Mutant 2.0. Positive $\Delta\Delta G$ values resemble high stability of the mutated protein and vice versa. [13]

PANTHER

Protein analysis using Protein Analysis through Evolutionary Relationship (PANTHER) tool organize proteins with families and subfamilies of evolutionarily related proteins relying on function, pathway or biological process of the proteins [14]. To estimate whether nsSNPs have and functional influence on the protein, PANTHER uses subPSEC score, which is calculated according to evolutionarily related proteins. SubPSEC score of less than -3 is predicted to be deleterious [15].

HOPE

Have your Protein Defined (HOPE) tool was used to assess the structural impact of rs111418524, rs41278081 on transferrin [16]. These approaches analyzed the impact of the mutation, and the findings showed a structural impact. This report also shows equivalence to contacts such as metal, DNA, hydrogen bonds, and ionic interactions and evaluates whether a mutation impacts an essential contact, structural areas together with motifs, domains, and trans-membrane domains. These tools assure that the foremost dependable method to acquire data and facts are used, i.e., data regarding the "actual protein structure" that provide annotated information in UniProtKB that is utilized by prediction with DAS servers [17].

RESULTS AND DISCUSSION

rs111418524, rs41278081 was submitted to different software, the SIFT **SIFT** program to check for the tolerance index (TI), **PolyPhen-2** to examine the possible impact of an amino acid substitution on the structure and function of a protein. **I-Mutant 2** to test for the stability of the protein and finally submitted to **PANTHER**, to confirm ^{the alignment of evolutionarily related} proteins.

As shown in Table 1, SIFT score was 0.00, 0.00 respectively, indicating in-tolerant (damage), PolyPhen2 score -1.000,-0.999 representing probably damaging the Protein function, I-Mutant 2 $\Delta\Delta G$ is -2.57, -2.48 this negative $\Delta\Delta G$ value indicate the low stability of the mutated ANXA5, and finally, SubPSEC score of PANTHER is 0.542, 0.543 which indicates a deleterious effect on protein function.

TABLE: I, RESULTS OF SNPS FROM DIFFERENT COMPUTATIONAL TOOLS.

GENE	Rs ID	AMINO ACID POSITION	SIFT	POLYPHEN	I-MUTANT 2.0	PANTHER
ANXA5	Rs111418524	S246G	0.00	-1.000	-2.57	0.542
ANXA5	Rs41278081	I81T	0.00	-0.999	-2.48	0.543

To find out possible impact of these mutations on ANXA5 structure, how your Protein Explained (HOPE) tool was used. This tool was developed at the center for molecular and biomolecular informatics (CMBI), department of bioinformatics, Radboud University. HOPE evaluates the effect of the mutation on the following features: the location and the interaction of the mutated residue, in addition, whether Variants in this residue, is common in other homologous proteins.

The Mutant protein structure depicted in Fig: 1 in cartoon form. The variation between wild and mutant protein structures are shown in figure 2 & 3. **rs111418524** encoded as missense mutation of a serine into a glycine at position 246, each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties. The mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions. (Fig: 4) Glycine is very flexible and can disturb the required rigidity of the protein at this position. The mutated residue is located in a domain that is important for binding of other molecules. The mutation might disturb the interaction between these two domains and as such affect the function of the protein.

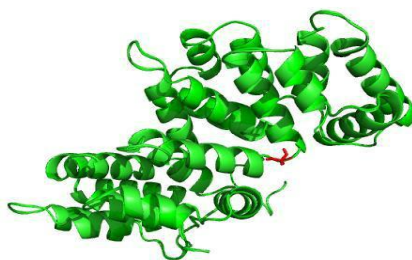


Fig: 1, Overview of the Protein (Mutant) In Ribbon-Presentation.
The protein is in green color, the side chain of the mutated residue is in red color.

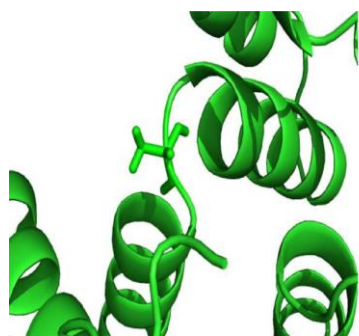
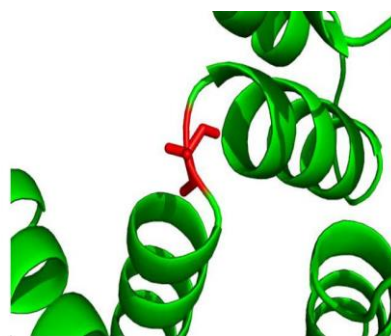


Fig: 2 Wild-Type Protein



Figs: 3 Mutant Type Proteins

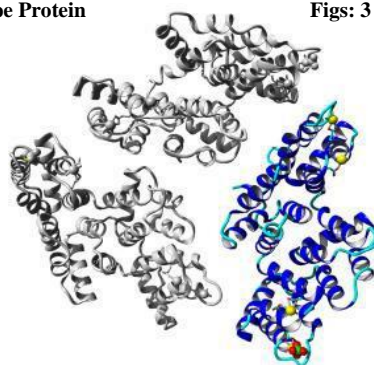


Fig: 4, Size of the Wild and Mutant Proteins

The wild-type and mutant amino acids differ in size. The mutant residue is smaller than the wild-type residue. This will cause a possible loss of external interactions.

The Mutant protein structure depicted in Fig: 5 in cartoon form. The structure of wild-type and mutant proteins are shown in figure 6 & 7. **rs41278081** encoded as missense mutation of an isoleucine into a threonine at position 81, each amino acid has its own specific size, charge, and hydrophobicity-value. The mutant residue is smaller than the wild-type residue. (Fig: 8) The wild-type residue is more hydrophobic than the mutant residue. The mutation will cause loss of hydrophobic interactions in the core of the protein.

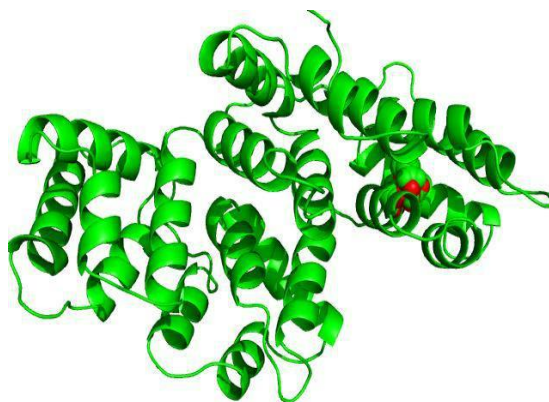


Fig: 5 Overview of the protein (Mutant) in ribbon-presentation.
The protein is in green color, the side chain of the mutated residue is in red color.

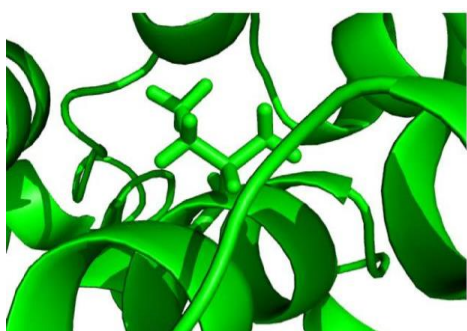
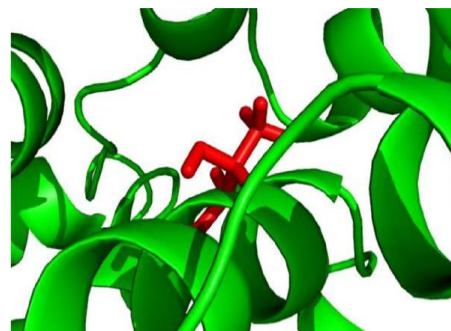


Fig: 6 Wild-Type Protein



Figs: 7 Mutant Type Protein

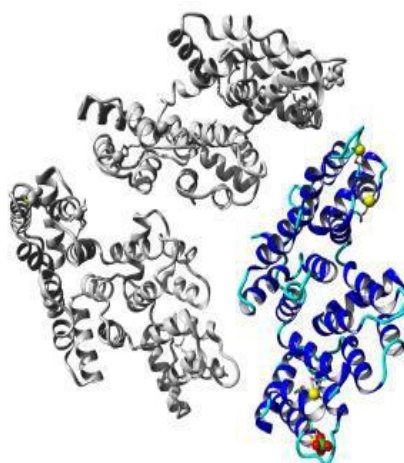


Fig: 8, Size of the Wild And Mutant Proteins

The wild-type and mutant amino acids differ in size. The mutant residue is smaller than the wild-type residue. The mutation will cause an empty space in the core of the protein

CONCLUSION

The present study offers an insight into the genotype–phenotype association SNPs associated with ANXA5 gene. Our study identified two pathogenic SNPs: rs111418524(S246G) and rs41278081(I81T) in ANXA5. The analysis of SIFT, PolyPhen, Mutant-I, Panther, and Hope has revealed their respective major consequences on the mutated ANXA5 gene, and also revealed their plausible malfunctioning mechanism via their structural destabilization. Compared to the wild type, all the selected mutations were altering the structural behavior of the protein, however, S246G (rs111418524) predicted to cause the

most damage to the protein's structural features followed by rs41278081 (I81T). Overall, the present computational approach will provide a comprehensive view on destabilizing mechanisms of Annexin A5 protein SNPs in ANXA5. The knowledge thus acquired through this present study is expected to help in prioritizing the important SNPs to be selected for further wet lab evaluations, and is of high value especially when designing huge population-based genotyping studies.

ACKNOWLEDGMENTS

The authors gratefully acknowledge to the UGC-MANF for providing the Fellowship for carrying out the Research work and to the Dept. of Biotechnology, Acharya Nagarjuna University for providing the facilities to carry out this work.

REFERENCE

1. B. Straypedersen, S. Straypedersen, "Etiologic factors and subsequent reproductive performance in 195 couples with a prior history of habitual abortion," *Am J Obstet Gynecol*, 148:140-146, 1984.
2. World Health Organisation. Recommended definitions, terminology, and format for statistical tables related to perinatal period. *Acta Obstet Gynecol Scand* 1977; 56:247-253.
3. Bettina Toth, Udo Jeschke, Nina Rogenhofer, Christoph Scholz, Wolfgang Würfel, J. Christian, Thaler, Antonis Makrigiannakis, "Recurrent miscarriage: current concepts in diagnosis and treatment", *Journal of Reproductive Immunology*, JRI-1910; No. of Pages 8, 2010.
4. K. H. Baek, R. M. Bertina, B.P. Koeleman, T. Koster, F.R. Rosendaal, Driven. "Aberrant gene expression associated with recurrent pregnancy loss," *Mol Hum Reprod*, 10(5): 291297. 2004.
5. M. A. Carvalho, S.M. Marsillac, R. Karchin, "Determination of cancer risk associated with germ line BRCA1 missense variants by functional analysis," *Cancer Research*, vol. 67, no. 4, pp. 1494-1501, 2007.
6. M. Carvalho, M. A. Pino, R. Karchin, "Analysis of a set of missense, frameshift, and in-frame deletion variants of BRCA1," *Mutation Research*, vol 660, no. 1-2, pp. 1-11, 2009.
7. D. E. Goldgar, D.F. Easton, A. M. Deffenbaugh, A. N. A. Monteiro, S.V. Tavtigian and F.J. Couch, "Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2," *American Journal of Human Genetics*, vol. 75, no. pp. 535- 544, 2004.
8. R. Karchin, "Next generation tools for the annotation of human SNPs," *Briefings in Bioinformatics*, vol. 10, no. 1, pp. 35-52, 2009.
9. P. C. Ng, S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function", *Nucleic Acids Res*, 31(13), 3812-3814, 2003.
10. S. F. Altschul, T.L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, ". Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res*, 25(17), 3389-3402, 1997.
11. I. A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, S. R. Sunyaev, "A method and server for predicting damaging missense mutations," *Nat Methods*, 7(4), 248-249, 2010.
12. I. Adzhubei, D. M. Jordan, S. R. Sunyaev, "Predicting the functional effect of human missense mutations using PolyPhen-2," *Curr Protoc Hum Genet*, Chapter 7, Unit-720, 2013.
13. K.A. Bava, M. M. Gromiha, H. Uedaira, K. Kitajima, A. Sarai, "ProTherm, version 4.0: thermodynamic database for proteins and mutants," *Nucleic Acids Res*, 32(Database issue), D120-121, 2004.
14. H. Mi, A. Muruganujan, P. D. Thomas, "PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees," *Nucleic Acids Res*, 41(Database issue), D377-386, 2013.
15. P. D. Thomas, M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, A. Narechania, "PANTHER: a library of protein families and subfamilies indexed by function. [Comparative Study]," *Genome Res*, 13(9), 2129-2141, 2003.
16. H. Venselaar, T. A. Te Beek, R. K. Kuipers, M. L. Hekkelman, G. Vriend, "Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces," *BMC Bioinformatics*, 11, 548, 2010.
17. Prlic, T. A. Down, E. Kulesha, R. D.Finn, A. Kahari, T. J. Hubbard, "Integrating sequence and structural biology with DAS," *BMC Bioinformatics*, 8, 333, 2007.