Journal of Carcinogenesis

Original Article

Exploration of structural stability in deleterious nsSNPs of the *XPA* gene: A molecular dynamics approach

NagaSundaram N., George Priya Doss C.[1]*

School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu, [1]Center for Nanobiotechnology, Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu, India

E-mail: georgecp77@yahoo.co.in
*Corresponding author

Published: 30 November, 2011
Journal of Carcinogenesis 2011, 10:26
This article is available from: http://www.carcinogenesis.com/content/10/1/26

Received: 27 August, 2011
Accepted: 20 October, 2011

Abstract

Background: Distinguishing the deleterious from the massive number of non-functional nsSNPs that occur within a single genome is a considerable challenge in mutation research. In this approach, we have used the existing *in silico* methods to explore the mutation-structure-function relationship in the *XPA* gene. **Materials and Methods:** We used the Sorting Intolerant From Tolerant (SIFT), Polymorphism Phenotyping (PolyPhen), I-Mutant 2.0, and the Protein Analysis THrough Evolutionary Relationships methods to predict the effects of deleterious nsSNPs on protein function and evaluated the impact of mutation on protein stability by Molecular Dynamics simulations. **Results:** By comparing the scores of all the four *in silico* methods, nsSNP with an ID rs104894131 at position C108F was predicted to be highly deleterious. We extended our Molecular dynamics approach to gain insight into the impact of this non-synonymous polymorphism on structural changes that may affect the activity of the *XPA* gene. **Conclusion:** Based on the *in silico* methods score, potential energy, root-mean-square deviation, and root-mean-square fluctuation, we predict that deleterious nsSNP at position C108F would play a significant role in causing disease by the *XPA* gene. Our approach would present the application of *in silico* tools in understanding the functional variation from the perspective of structure, evolution, and phenotype.

Keywords: I-Mutant, molecular dynamics, PANTHER, polyphen, SIFT, *XPA*

INTRODUCTION

Common polymorphisms in *Nucleotide Excision Repair (NER)* genes have been studied extensively in terms of their association with the risk of various cancers.[1] The rationale behind these gene-cancer risk associations is that these genetic variants may result in alterations in phenotypes (i.e., DNA repair capacity). In humans, defects in NER, which are caused by mutations in the *NER* genes, are associated with a rare, cancer-prone syndrome called *xeroderma pigmentosum* (XP). Key proteins in the transcription-coupled *nucleotide excision repair (NER)* pathway include the Xeroderma pigmentosum group A (*XPA*). Single Nucleotide Polymorphisms (SNPs) are the most common and simplest type of genetic variations. SNPs that change the encoded amino acids are called non-synonymous single nucleotide polymorphisms (nsSNPs) that result in amino acid substitution and are often associated with disease. They are believed to have the greatest impact on protein function because the mutations of the encoded

Access this article online

Quick Response Code:

Website: www.carcinogenesis.com

DOI: 10.4103/1477-3163.90441

amino acids can have a deleterious effect on the structure and / or function of the proteins.[2] The identification of the nsSNPs responsible for specific phenotype variations requires multiple testing of hundreds or thousands of SNPs in the candidate genes. *In vivo* studies on the function of nsSNPs have found that mutations in *XPA* are associated with various cancers.[3-5] Experimentally based approaches are labor-intensive and time-consuming, and at the structural level, often nearly impossible, especially in cases where there are several mutations causing the disease. By contrast, *in silico* methods have high priority in characterizing variants, because they can be employed in the systematic screening of representatives of human populations, for variations. *In silico* methods use alternative classification methods (sequence and / or structure) to predict the deleterious amino-acid substitutions that impact protein structure and activity on a larger scale. In this context, we used the Sorting Intolerant From Tolerant (SIFT), Polymorphism Phenotyping (PolyPhen), I-Mutant 2.0, and PANTHER methods. To develop a coherent approach in the structural analysis, we applied Molecular Dynamics (MD) simulations using the Groningen Machine for Chemical Simulations (GROMACS), as a structural perspective to investigate protein stability in both native and mutant models. We compared the dynamic behavior of the native and mutant models by analyzing the potential energy, root-mean-square deviation (RMSD), and root-mean-square fluctuation (RMSF) of the C108F mutations.

MATERIALS AND METHODS

For our analysis, SNPs and their related information on the *XPA* gene was retrieved from the National Center for Biotechnology Information and SWISS-Prot databases. The 3D structure of the *XPA* gene, 1*XPA,* was obtained from the Protein Data Bank. Comparative *in silico* analyses were performed to identify the deleterious nsSNPs in the *XPA* gene. We used a sequence-based approach SIFT, which predicted whether an amino acid substitution affects the protein function, based on sequence homology and the physical properties of amino acids.[6] A SIFT score ≤ 0.05 indicates that the amino acid substitution is intolerant or deleterious, whereas, a score ≥ 0.05 predicts to be tolerant. PolyPhen uses sequence, phylogenetics, and structural information for characterizing the deleterious substitution. A Position Specific Independent Count (PSIC) score difference of 1.5 and above is considered to be damaging.[7] PANTHER uses HMM-based statistical modeling methods and multiple sequence alignments to perform the evolutionary analysis of coding nsSNPs. PANTHER estimates the likelihood of a particular nsSNP causing a functional impact on the protein.[8] The PANTHER subPSEC scores vary from 0 (neutral) to about -10 (most likely to be deleterious). Protein sequences having a subPSEC value ≤ -3 are said to be deleterious. I-Mutant 2.0[9] is a support vector–based method, for the automatic prediction of protein stability changes on single point mutations. The output file shows the predicted free energy change (DDG), which is calculated from the unfolding Gibbs free energy change of the mutated protein, minus the unfolding free energy value of the native protein (Kcal / mol). DDG > 0 means that the mutated protein has high stability and vice versa.

A molecular dynamics analysis was performed based on the crystal structure of the protein for evaluating the structural stability of native and mutant proteins. Information about mapping the nsSNPs in the protein structure was obtained from dbSNP. The *XPA* protein coordinates were obtained from the Protein Data Bank with entry 1*XPA* solved by nuclear magnetic resonance (NMR).[10] Mutation was introduced using the SWISS-PDB viewer. The missing hydrogen atoms and broken chain were added using Visual Molecular Dynamics.[11] Energy-minimized structures of the native and mutant *XPA* proteins were used as a starting point for molecular dynamics simulation analysis. Molecular mechanics potential energy minimizations and molecular dynamics simulations were carried out using the program package GROMACS 4.0.5.[12,13] Force field GROMOS96[14] 43a1 was used in all molecular dynamics simulations. A periodic boundary condition was applied, and the pressure and temperature were kept constant in the system. The temperature was kept constant by using the Berendsen algorithm[15] with a coupling time of 0.2 ps. The protein was then solvated into a 0.9 Å cubic box of water with a Simple Point Charge model.[16] The system was then neutralized by adding chlorine ions Cl^- around the molecule, with the use of gen-ion command. In this step, six ions were added to both native and mutant structures. Finally, random water molecules were substituted with six sodium ions, in order to obtain neutralized system. Van der Waals interaction cutoff was set to 1.4 Å, and long-range electrostatic forces were computed using the particle-mesh Ewald summation method. The trajectories at 300 K for 4 ns were considered to be the most probable structures under physiological conditions, and were analyzed in detail. The RMSD of the backbone atoms, the RMSF of carbon-alpha carbon, and the potential energy projection of the protein in the phase space of the system were plotted using the GRaphing, Advanced Computation and Exploration (GRACE) program.

RESULTS AND DISCUSSIONS

Seven nsSNPs of the *XPA* gene investigated in this analysis were retrieved from the dbSNP database. The functional

impact of these nsSNPs could be assessed by evaluating the importance of the amino acids they affected. We employed four widely-used computational tools for determining the functional significance of the nsSNPs in this analysis. SIFT predicted whether an amino acid substitution affected the protein function based on sequence homology and the physical properties of the amino acid**.** nsSNP with an ID rs104894131, which showed a SIFT score of 0.00, was predicted to be highly deleterious. Polyphen predicted the effect of nsSNP on amino acid residues by a structure-based approach. All protein sequences submitted to SIFT were also submitted to PolyPhen. Two nsSNPs with IDs rs104894131 and rs57519506 exhibited Polyphen scores of 3.08 and 1.650, respectively. They were predicted to be probably damaging the protein structure and function. We further analyzed seven nsSNPs using PANTHER, for validating their impact on protein function upon a single point mutation. Out of these seven nsSNPs taken for our analysis, only two nsSNPs with IDs rs104894131 and rs75031098 showed a subPSEC score of - 6.5519 and - 3.07023, and were designated as deleterious. The protein stability change due to a single point mutation was predicted using a support vector machine-based tool, I-Mutant 2.0. Out of seven nsSNPs, four nsSNPs with IDs rs104894131, rs75031098, rs3176749, and rs1805160 showed DDG values of -1.48, -1.75,-0.40, and -0.58 Kcal / mol, and were considered to be the least stable and deleterious. We found that there was limited concordance among these *in silico* methods. Most of these differences were likely the result of each method requiring a sufficient number and diversity of aligned sequences in order to make a prediction and each method using a different set of sequences and alignments. By comparing the scores of all the four methods used in this analysis, nsSNP with an ID rs104894131 was predicted to be highly deleterious and functionally significant. Hence, we concluded that this particular substitution should be considered as an important candidate in causing disease related to *XPA* gene malfunction. For the intensive computational studies, the molecular dynamics simulations were carried out in the nsSNP with an ID rs104894131, at position C108F of the *XPA* gene. Molecular dynamics simulation may be useful to gain insight into the impact of non-synonymous polymorphisms on structural changes that may affect the activity of the *XPA* gene.

Potential energy minimizations were performed on each of the initial systems. Potential minimized energy values of native and mutant structures were -6.8477925 and -6.8200406, respectively. An energy value reflected that the mutant protein had obtained less energy value, when compared to the native protein. This revealed that the mutation at the one hundred and eighth position (C→F)

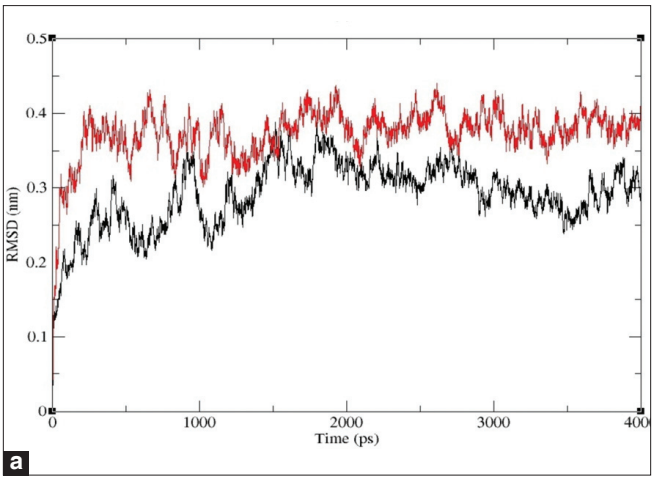

Figure 1a: RMSD of the backbone atoms in 1XPA protein. Black and Red lines indicate RMSD of the native and mutant structures. The ordinate is RMSD (a), and the abscissa is time (ps)

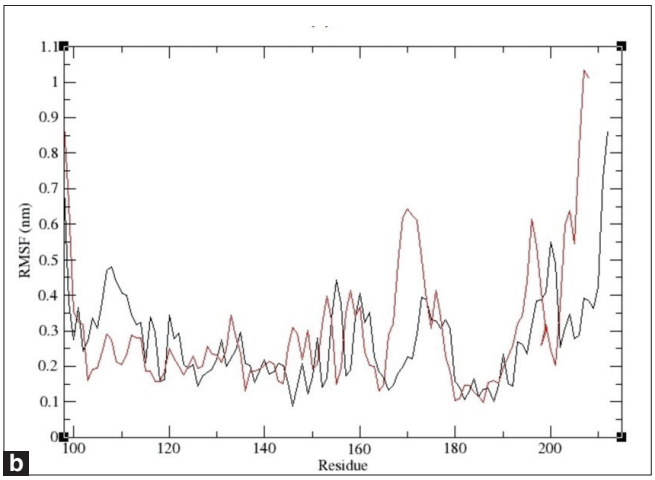

Figure 1b: RMSF of the backbone carbon alpha over the entire simulation. Black and Red lines indicate RMSF of the native and mutant structures. The ordinate is RMSF (nm), and the abscissa is residues

of the *XPA* gene affected the stability of the protein. The RMSD value of the backbone atoms of the entire protein was calculated from the trajectory data. The nsSNP at the one hundred and eighth position (C→F) exhibited a notable deviation from the starting point itself. The deviation of RMSD for the native protein was initiated around ~0.2 Å and mutant protein in the region of ~0.3 Å within the range of ~100 ps. At ranges of 1000 ps, 1500 ps, and 2000 ps, as in Figure 1a, a similar fashion of deviation is seen in the same region of ~ 0.35 Å. Native protein showed a deviation in the minimum region of ~0.2 Å and maximum of ~0.35 Å within the range of 1000 ps to 4000 ps. Although the mutant protein showed a deviation in the minimum region of ~0.3 Å and maximum of ~0.4 Å in the same region range. RMSD in the native protein was entirely deviated from the mutant structure in the whole simulation. This constant range of

deviation in the native and mutant structure reflected that the mutation affected the effects of the dynamic behavior of the residues. The stabilizing effects of the protein are reflected by the RMSF values. The RMSF value of the alpha carbon of each amino acid residue was calculated from the trajectory data of the native and mutant proteins [Figure 1b]. nsSNP with an ID rs104894131, at position C108F on the *XPA* gene, exhibited great fluctuations of the entire protein sequence itself. Notably an initial rise in RMSF was observed between the native and mutant structures in the residue position ~105 to 115. These results confirmed that our prediction that deleterious nsSNP at position C108F would play a significant role in causing disease by the *XPA* gene. Our approach would present the application of *in silico* tools in understanding the functional variation from the perspective of structure, evolution, and phenotype.

ACKNOWLEDGMENTS

The authors thank the management of the VIT University for providing the facilities to carry out this study.

REFERENCES

1. Lin J, Swan GE, Shields PG, Benowitz NL, Gu J, Amos CI, *et al*. Mutagen sensitivity and genetic variants in nucleotide excision repair pathway: genotype-phenotype correlation. Cancer Epidemiol Biomarkers Prev 2007;16:2065-71.
2. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 2001;307:683-706.
3. Wu X, Zhao H, Wei Q, Amos CI, Zhang K, Guo Z, *et al*. *XPA* polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. Carcinogenesis 2003;24:505-9.
4. Park JY, Park SH, Choi JE, Lee SY, Jeon HS, Cha SI, *et al*. Polymorphisms of the DNA repair gene xeroderma pigmentosum group A and risk of primary lung cancer. Cancer Epidemiol Biomarkers Prev 2002;11:993-7.
5. Lockett KL, Snowhite IV, Hu JJ. Nucleotide-excision repair and prostate cancer risk. Cancer Lett 2005;220:125-35.
6. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.
7. 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;7:248-9.
8. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, *et al*. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 2003;13:2129-41.
9. Capriotti E, Fariselli P, Casadio R. I-Mutant 2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33:306-10.
10. Ikegami T, Kuraoka I, Saijo M, Kodo N, Kyogoku Y, Morikawa K, *et al*. Solution structure of the DNA- and RPA-binding domain of the human repair factor *XPA*. Nat Struct Biol 1998;5:701-6.
11. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:27-8,33-8.
12. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 2008;4:435-47.
13. Kutzner C, van der Spoel D, Fechner M, Lindahl E, Schmitt UW, de Groot BL, *et al*. Speeding up parallel GROMACS on high-latency networks. J Comput Chem 2007;28:2075-84.
14. Van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Krüger P, Mark AE, *et al*. Biomolecular Simulation: The GROMOS96 Manual and User Guide; vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v: Zurich, Groningen, 1996.
15. Berendsen HJ, Postma JP, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684-90.
16. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926.

How to cite this article: NagaSundaram N, Priya Doss CG. Exploration of structural stability in deleterious nsSNPs of the *XPA* gene: A molecular dynamics approach. J Carcinog 2011;10:26.

AUTHOR'S PROFILE

Mr. NagaSundaram N, School of Biosciences and Technology, VIT University, Vellore 632014, Tamil Nadu, India

Dr. George Priya Doss C, Center for Nanobiotechnology; Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore 632014, Tamil Nadu